Guidelines for Health Technology Assessment in Thailand (Second Edition)
# Contents

## A. Special Article

- **Where are the Limits of Cost-Effectiveness Analysis and Health Technology Assessment?**  
  Anthony J. Culyer  
  
  - **Health Technology Assessment in Developing the National List of Essential Medicines in Thailand**  
    Visanu Thamlikitkul  
    
    - Page: S1
    - Page: S3

## B. Original Article

- **Guidelines for Health Technology Assessment in Thailand (Second Edition)-The Development Process**  
  Usa Chaikledkaew, Kankamon Kittrongsiri  
  
  - Page: S4
- **Defining the Scope of Health Technology Assessment and Types of Health Economic Evaluation**  
  Pimwara Tanvejsilp, Surachat Ngorsuraches  
  
  - Page: S10
- **Measurement of Costs for Health Economic Evaluation**  
  Arthorn Riewpaiboon  
  
  - Page: S17
- **Measurement of Health Outcomes**  
  Montarat Thavorncharoensap  
  
  - Page: S27
- **Systematic Review and Network Meta-Analysis in Health Technology Assessment**  
  Nathorn Chaiyakanapruk, Surasak Saokaew, Rosarin Sruamsiri, Piyameth Dilokthornsakul  
  
  - Page: S33
- **Measurement of Utility**  
  Montarat Thavorncharoensap  
  
  - Page: S43
- **Handling Time in Economic Evaluation Studies**  
  Unchalee Permsuswan, Kansinee Guntawongwan, Piyaluk Buddhawongsa  
  
  - Page: S50
- **Sensitivity Analysis for Handling Uncertainty in an Economic Evaluation**  
  Supon Limwattananan  
  
  - Page: S59
- **Budget Impact Analysis**  
  Pattara Leelahavarong  
  
  - Page: S65
- **Presentation of Economic Evaluation Results**  
  Usa Chaikledkaew  
  
  - Page: S72
- **Social and Ethical Analysis in Health Technology Assessment**  
  Sripen Tantivess  
  
  - Page: S81
- **A Way Forward for the Evaluation of Health Technologies for Infectious Diseases**  
  Pritaporn Kingkaew  
  
  - Page: S87
- **Economic Evaluation of Screening for Disease**  
  Naiyana Praditsithikorn  
  
  - Page: S94
- **The Economic Evaluation of Medical Devices: Challenges**  
  Pritaporn Kingkaew, Yot Teerawattananon  
  
  - Page: S102
- **Constructing a State-Transition Model for an Economic Evaluation of Cancer Treatments**  
  Chulaporn Limwattananon, Supon Limwattananon  
  
  - Page: S108
- **Quality Assessment of Health Economic Evaluation**  
  Usa Chaikledkaew, Kankamon Kittrongsiri  
  
  - Page: S113
- **Application of HTA Research on Policy Decision-Making**  
  Sitaporn Youngkong  
  
  - Page: S119
- **Standard Cost Lists for Health Economic Evaluation in Thailand**  
  Arthorn Riewpaiboon  
  
  - Page: S127
Executive Committee 2012 - 2013

The Council of the Medical Association

President: Wonchart Subhachaturas
President Elect: Prasert Sarnvivad
Vice-President: Theerachai Chanthanarojsiri
Secretary General: Saranatra Waikakul
Deputy Secretary-General: Chanrit Lawthaweesawat
Treasurer: Paisal Chantarapitak
House Master: Sawat Takearngej

Chairman

Publication: Amorn Leelarasamee
Welfare Section: Nithiwat Gijsriurai
Scientific Section: Somkiat Wattanasirichaigoon
Medical Education: Kidaphol Wadhanakul
Medical Ethics: Ronnachai Kongsakon
International Relation: Nopadol Wora-Urai
Public Relation: Ittaporn Kanacharoen
Registration: Somchai Thepcharoenmirund
Special Affair: Juvady Leawpairat

Member of Council

Adisorn Patradul Pakpoom Supiyaphun Somsak Lolekha
Amnaj Kussalanan Parnrudee Manomaipiboon Soottiporn Chittmitrapap
Chalerm Harphanich Permyot Kosolbhand Sukajan Pongprapai
Chumsak Pruksapong Pinit Hirunyachote Thongchai Triviboonvanich
Kachit Choopanya Pirapong Saichuea Varaphan Unchak
Kammant Phanthumchinda Pongsak Wannakrairot Vira Khuangtsirikul
Kamol Veerapradist Sakchai Vongkittirux Wanna Sirojjanakul
Orawan Kiriwat Somboon Kunathikom Yothin Benjawung
INTERNATIONAL EDITORIAL BOARD
Akihiko Okuyama, Japan
Alfred J Tria, United States
Andrea Todisco, United States
Barbara E Murray, United States
Didier Raoult, France
Elizabeth Frost, United States
Frans MJ Debruyne, Netherland
Hextan Ngan, Hong Kong
James Scheiman, United States
John Howard Scholefield, United Kingdom
Kenrad E Nelson, United States
Louis CK Low, Hong Kong
Richard J Martin, United States
Robin Williamson, United Kingdom
Tippavan Nagachinta, United States
Yuji Murata, Japan

SENIOR ADVISORY BOARD
Arun Pausawasdi
Athisit Vejajiva
Charas Suwanwela
Ouichai Pleangprasit
Prasert Thongcharoen
Somchai Boornkitti
Somrsi Pausawasdi
Somwang Danchaiyijitr
Sira Bunyaratavci
Sukhit Phaosavasdi
Visith Sitprija
EDITOR - IN - CHIEF
Amorn Leelarasamee
CO - EDITORS
Anek Aribarg
Apichard Sukonthasarn
Kamheang Chaturachinda
Polasak Jeeravipoolvan
Prasit Waranapa
Prapat Singhasivanon
Ronmachai Kongsaikon
Somwang Danchaiyijitr
Sunthon Horpaopan
Supasit Pannarunothai
Suthipun Jitpimolmard
Verapol Chandeying

EDITORIAL BOARD
Amnuay Cutchavaree
Anupong Chitwarakorn
Apichana Kovindha
Apichart Singalavanijia
Bangchong Mahaisavijira
Boonchai Uerpairojkit
Chatlert Pongchayakul
Chaweewan Bunnag
Jariya Lertakamanee
Kannika Phonphurtkul
Kobchiti Limpaphayom
Lucksana Pochnugool
Nopadol Wora-Urai
Nalinee Aswapokee
Nanta Maraneta
Nualanong Visitsunthorn
Niphon Poungvarin
Ounjai Kor-anantakul
Orasa Chawalparit
Pairat Tepmongkol
Paksit Vichayanond
Pansak Laksanabunsong
Panupan Songcharoen
Paramee Thongsuksai
Pipop Jirapinyo
Pisaln Mairiang
Ploenchan Chetchetisakd
Pornswan Wasant
Rajata Rajatanavin
Saranatra Waikakul
Sonrat Charuluxananan
Somdak Tiamkao
Supakorn Rojananin
Surachai Potchasasupawan
Surapol Issaragrisil
Surasak Taneepanichskul
Suthat Fucharoen
Suvit Srussadaporn
Tanapon Maipang
Thep Himathongkam
Thewarug Werawataganon
Ungkab Prakanrattana
Vajarabongsa Bhudhisawasdi
Vichai Chirarakon
Virat Kirtsreesakul
Visit Prasartwanakit
Vithaya Sridama
Vitoon Chinsangwatanakul
Vorapa Suwanchinda
Voraphong Phupong
Wannee Nitiyanant
Wichien Laohacharoenbomat
Wiroon Laupattarakasem
Yong Poovorawan
Yosanan Yospaiboon
Yuen Tannirandorn

ETHICAL SECTION EDITORS
Chumsak Pruksapong
Sukhit Phaosavasdi
Surasak Taneepanichskul
Yuen Tannirandorn

MANAGEMENT BOARD
Arun Pausawasdi
Athasit Vejajiva
Charas Suwanwela
Ouichai Pleangprasit
Prasert Thongcharoen
Somchai Boornkitti
Somrsi Pausawasdi
Somwang Danchaiyijitr
Sira Bunyaratavci
Sukhit Phaosavasdi
Visith Sitprija

THE MEDICAL ASSOCIATION OF THAILAND
Office: 4th floor, Royal Golden Jubilee Building, 2 Soi Soonvijai, New Petchburi Road, Huaykwang, Bangkok 10310, Thailand
Phone: (66-2) 314-4333, (66-2) 318-8170, (66-2) 716-6102, (66-2) 716-6962 press 0
Fax: (66-2) 314-6305 e-mail: jmedassocthai@yahoo.com
Website: http://www.mat.or.th e-Journal: http://www.jmatonline.com

THE OFFICIAL JOURNAL OF
The Medical Association of Thailand
The Royal Thai College of Obstetricians and Gynaecologists
The Royal College of Physicians of Thailand
Faculty of Medicine, Prince of Songkla University
Commission on Higher Education, Bureau of Personnel Administration and Development
Faculty of Medicine, Naresuan University
Faculty of Medicine, Khon Kaen University
College of Public Health Sciences, Chulalongkorn University

Frans MJ Debruyne, Netherland
Hextan Ngan, Hong Kong
James Scheiman, United States
John Howard Scholefield, United Kingdom
Kenrad E Nelson, United States
Louis CK Low, Hong Kong
Richard J Martin, United States
Robin Williamson, United Kingdom
Tippavan Nagachinta, United States
Yuji Murata, Japan

Frans MJ Debruyne, Netherland
Hextan Ngan, Hong Kong
James Scheiman, United States
John Howard Scholefield, United Kingdom
Kenrad E Nelson, United States
Louis CK Low, Hong Kong
Richard J Martin, United States
Robin Williamson, United Kingdom
Tippavan Nagachinta, United States
Yuji Murata, Japan
Guidelines for Health Technology Assessment in Thailand (Second Edition)

The publication of the first edition of HTA guidelines has been recognized as a significant step forward in improving HTA research and development and thus policy-making. They have received widespread acceptance in the scientific and policy making communities in Thailand, and have been approved by the Subcommittee for Development of the National List of Essential Drugs and the Subcommittee for Development of the Health Benefit Package and Service Delivery of the National Health Security Office. However, it is important to note that guidelines such as these should always be regarded as dynamic tools that will require refinement over time and should be adapted according to the changing nature of Thailand’s healthcare context. Therefore, the development of this second edition of HTA guidelines for Thailand builds on the success of the first edition, while attempting to address some of the identified limitations of the first edition and reflect the changes that the health care and policy contexts have undergone in the intervening years.

As editors of this special volume, we wish to acknowledge all individuals and related organisations that contributed throughout the process for guideline development. We would like to cordially thank Health Intervention and Technology Assessment Program (HITAP), the body that serves as the coordinator of Thai HTA guidelines, as well as the second HTA guideline development working group consisting of the experts from both academic and research institutions across the country for their invaluable inputs and significant support. 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.

Yot Teerawattananon MD, PhD
Program Leader
Health Intervention and Technology Assessment Program (HITAP)

Usu Chaikledkaew BSc(Pharm), MA, PhD
Associate Professor
Faculty of Pharmacy,
Mahidol University
HEALTH INTERVENTION AND TECHNOLOGY ASSESSMENT PROGRAM

Office: 6th Floor, 6th Building, Department of Health, Ministry of Public Health, Tiwanon Road, Nonthaburi 11000, Thailand.
Phone: (66-2) 590-4549, (66-2) 590-4373  Fax: (66-2) 590-4369  Website: www.hitap.net

EDITORIAL BOARD OF THIS SUPPLEMENT ISSUE
Journal of the Medical Association of Thailand (Vol.97 Suppl.5 2014)

Editor
Usa Chaikledkaew
Yot Teerawattananon

Review
Arthorn Riewpaiboon  Pattara Leelahavarong  Sripen Tantivess
Chulaporn Limwattananon  Pimwara Tanvejsilp  Supon Limwattananon
Kankamon Kitttrongsiri  Piyaluk Buddhawongsa  Surachat Ngorsuraches
Kansinee Guntawongwan  Piyaemeth Dilokthornsakul  Surasak Saokaew
Montarat Thavorncharoensap  Pritaporn Kingkaew  Unchalee Permsuwan
Naiyana Praditsitthikorn  Rosarin Sruamsiri  Usa Chaikledkaew
Nathorn Chaiyakunapruk  Sitaporn Youngkong  Yot Teerawattananon
A guide for authors

The Journal of Medical Association of Thailand strongly recommends to the authors of the manuscript to observe and follow the code on good publication practice produced by the Committee on Publication Ethics (http://www.publicationethics.org.uk/guidelines/), the uniform requirements for manuscript and potential conflicts of interest reporting produced by ICMJE (http://www.icmje.org/) and the EQUATOR network resource centre (http://www.equator-network.org/) for good research reporting.

Manuscript Preparation: The original typescript, in print and on data file diskettes (CD only) must be submitted, including copies of all illustrations, legends and references. All must be typed double-spaced on 8½" x 11" (A4) white bond paper. Margins must be at least 1 inch. Only a word processor should be used. Metric measurements are to be used throughout the text and illustrative materials. Generic names for drugs are to be used. Brand names may be included in parentheses.

Titles: The title must be short, specific, and clear. The title page should be typed on a separate sheet and include the full name, degree and academic affiliations of all authors.

Abstracts: Submit abstracts in duplicate, both in English and Thai, which must include a concise statement in not more than 200 words of what was done, what was found and what was concluded. If there is no Thai author, Thai abstract is not required.

Illustrations and Tables: Use only illustrations that clarify and augment the text. Submit illustrations in duplicate, unmounted, and untrimmed. High contrast, glossy-prints (not photocopies) are requested. Original drawings should be professionally drawn in black India ink. Figure number, name of senior author, and arrow indicating “top” should be typed on a gummed label and affixed to the back of each illustration. All lettering must be legible after reduction to column size. Each table should be typed double-spaced on a separate sheet, and should supplement rather than duplicate the text. They should be numbered consecutively as mentioned in the text and each must have a heading. Color illustrations will be published at the author’s expense.

References: References should be listed in consecutive order as they are cited in the text, not alphabetically. Once a reference is cited, all subsequent citations should be running in queue to the original number. All references must be cited in the text or tables. Unpublished data and personal communication are not accepted. Use the form of references and the titles of journals abbreviated, according to the style used in Index Medicus (the Vancouver system). If there are more than six authors, the first six authors are listed followed by “et al.”

Manuscript Acceptance: Manuscripts are received with the understanding that they are not under simultaneous consideration by another publication. Accepted manuscripts become the property of the Journal of the Medical Association of Thailand and must not be published in other journal.

Checklist before Manuscript Submission:
- Three copies of manuscript
- Three copies of abstract (Thai and English)
- Three copies of attached photographs, legends, and/or illustrations, clearly marked as described in the instructions
- One data file diskette or send file to jmedassocthai@yahoo.com
- The addresses, e-mail addresses, telephone numbers and fax numbers of the first and corresponding authors
- Other relevant documents such as potential conflicts of interest form, copy of EC approval for research form, trial registration number or document, etc.
Policy on Transparent Reporting of Health Research

To promote the transparent reporting of health research, the Journal of Medical Association of Thailand strongly recommends to the authors to observe and follow:

- The code on good publication practice produced by the Committee on Publication Ethics (http://www.publicationethics.org.uk/guidelines/)
- The uniform requirements for manuscript and potential conflict of interest reporting produced by ICMJE (http://www.icmje.org/)
- The uniform requirements for good research reporting by the EQUATOR network resource centre (http://www.equator-network.org/)

The report of a randomized controlled trial (RCT) should use the checklist and flow diagram as described in CONSORT statement 2010 as the minimum set of recommendations for reporting the RCT. The checklist items include the 25 items and pertain to the content of the Title, Abstract, Introduction, Methods, Results, Discussion, and Other information.

Similarly, we recommend to use the checklists for other types of studies such as systematic reviews (PRISMA), meta-analyses of observational studies (MOOSE), diagnostic accuracy studies (STARD), cohort, case control, cross-sectional studies (STROBE) and qualitative studies (COREQ or RATS [http://www.biomedcentral.com/info/ifora/rats]) when writing report using these studies. The authors can consult the EQUATOR network website for further information on the available reporting guidelines for health research. Though the implementation of this policy is rather new to Thai authors, the J Med Assoc Thai will strongly request potential authors to follow these guidelines as soon as possible.

Clinical Trial Registration Policy

The J Med Assoc Thai encourages potential Thai investigators who will conduct clinical trials and plan to publish their study results in J Med Assoc Thai, to register the clinical trials with an official trial registration organization before any participants of the trial are recruited. If you are not aware of any existing trial registration organizations or would like to seek for high quality clinical trial registers, please visit WHO Clinical Trial Search Portal at http://www.who.int/trialsearch/. For more information on the WHO International Clinical Trials Registry Platform go to http://www.who.int/ictrp. Alternatively, you may register the trial with the Thai Clinical Trials Registration (TCTR) at http://www.clinicaltrials.in.th. Now, the editor office starts to request for the registration document or number from the first or corresponding authors of the manuscript submitted to J Med Assoc Thai from August 2010. The authors should automatically show the registration document or number within or accompanying the submitted manuscript if this is available.

Research Ethics Review Committee (ERC or EC) Form

The first or corresponding author must submit one copy of the EC permission to research form with the signature for endorsement of its originality and accuracy.

Report of Potential Conflicts of Interest

All authors must disclose all relationships that could be viewed as potential conflicts of interest. Use the latest (2010) reporting form from the ICMJE (http://www.icmje.org/).

Plagiarism and Duplicate Publication

Plagiarized or duplicate manuscript/article will be rejected or retracted with penalty.

Copyright Notice

All materials or publications in the hard copy or showed on the website are protected by Medical Association of Thailand copyright and may not be reproduced, distributed, transmitted, displayed, or otherwise published without the prior written permission of the office of the Journal of Medical Association of Thailand.

However, members of the Medical Association of Thailand may download the .pdf file for personal, non-commercial, educational use only.

Any information in the hard copy or posted on the web is for informational purposes only. The Journal of Medical Association of Thailand is not responsible for the information in the article or the result of its practice.
Where are the Limits of Cost-Effectiveness Analysis and Health Technology Assessment?

Anthony J Culyer, CBE, BA, Hon DEcon, Hon FRCP, FRSA, FMedSci*

* Professor at the University of Toronto, Canada and University of York, England

The advent of a second edition of the Thai guidance on HTA, with its admirable stated aims of clarifying, updating and widening the scope of HTA in order to include budget impact analysis and assessments of social and ethical impact, seems a good time to ask what the limits of the scope of HTA truly are, and whether they are inherently conceptual in character or empirical and practical. One might think it a bit late to be reflecting thus after the revisions and all the brainstorming and consultations have been completed—rather like shutting the stable door after the horse has bolted. On the other hand, to have reached conclusions about some of the extended scope is not necessarily to have concluded any discussion of what the scope can or ought to be. In this short editorial I aim to set out a few ideas that may serve at least in part as an agenda for the third revision!

The authors have an excellent definition of HTA:

Health technology assessment is a [form of] policy research which integrates multidisciplinary fields in a systematic way in order to determine the effects on the use of health technology in short and long terms, direct and indirect effects, intentional or unintentional effects, effects of the development and diffusion of technology, and the group of related technologies and issues related to the application of technology[1].

The trouble with this definition is that it is a statement of aspiration rather than a statement of what HTA can or does do. There is nothing wrong with the aspiration. On the contrary there is a good deal that is right about it. However, HTA, as currently practiced, does not “integrate multidisciplinary fields” nor does it approach the issues in trying to do so in “a systematic way”. The list of short and long-term intentional and unintentional consequences is a very thinly populated list in all examples in practice.

My own view is that the reasons for this state of affairs are partly inherent and conceptual and partly empirical and practical. If the only difficulties were empirical and practical, I do not doubt that solutions would be found and the central issue would be one of assessing the value of the additional information. An appropriate way of thinking about those issues is readily available[2-4].

The inherent and conceptual difficulties have an altogether different character. The main problem here is that we do not have an all-encompassing theory of human welfare that integrates the various categories of effect that might properly be appraised. There are many lesser conceptual problems too, such as the metrics by which “more” or “less” of some of the entities of interest might be measured (“distributional fairness”? “fairness of what”-health, health care, social care?... “enhanced dignity?” “patient autonomy?” and so endlessly on). I do not wish to state that deriving measures that have construct validity and an appropriate degree of cardinality is impossible. For most of these items, however, it remains still to be done. But the prime difficulty lies in the combining. The economist’s idea of a social welfare function[5,6] is not quite what we seek, for its arguments are utilities and utilities derive from preferences. It is far from clear that the expanded set of variables that would meet the aspiration cited above can be limited to people’s preferences alone.

The evidence on the limited range of cost and benefit variables commonly embraced in HTA already poses considerable problems of interpretation: the evidence may be partial, it may be scientifically controversial, it often has high internal validity but low external validity, it may in some important respects be missing, it may not all be of comparable quality, some of it may be highly quantitative and other elements may be qualitative. Some may be based on observational...
studies with poor controls for confounding factors, some may be based on professional and probably biased opinion, much will have unknown and even unknowable biases.

How much more complex is the combining of a much wider set of desired elements. They will not combine themselves to produce health system guidance, instead, combining and interpreting them requires a deliberative process. A deliberative process is participative and often follows a period of consultation with relevant stakeholders. It entails both the eliciting and the combining of various types of evidence in order to reach an evidence-based judgment. There is little evidence on the effectiveness of deliberative processes, though there is much to be said in favor of them on grounds of principle. The design of a deliberative process is not neutral and may well influence the relative weights assigned to different types of evidence, thus influencing the extent to which the eventual guidance is “evidence-based”. Characteristics of a deliberative process likely to ensure evidence-based guidance include consultation with all parties affected by the outcome, fair representation of scientists and stakeholders, high-quality syntheses of the scientific evidence, and skillful chairing(7-10).

The way ahead thus seems to me to be one in which our prime focus should be less on refining and expanding our calculus of choice and more on the decision process and the design of processes that enable decision-makers to weigh up entities that are not only measured or indicated with varied precision and unknown biases, and not only incomplete and contestable, but that also belong in different layers of understanding and difficulty of comparison: variables arising from preferences, from notions of duty, from religious observance, from historical circumstance. Variables that may not be continuous but binary, like “good” and “bad”. The focus should not, moreover, be only on process. It should also be on the people who engage in the process. The processes we need are ones in which (unprejudiced) people can exercise an informed judgment. People do not come ready for these tasks. They need training in the testing and interpretation of information and in the exercise of good judgment(11).

The authors have hardly begun to address this agenda but, if progress can be made in these respects, then the boundaries of CEA and HTA will no longer be set by the limitations of the HTA algorithm but by the limitations of the human imagination. That is to say, they become virtually limitless!

References
1. Teerawattananon Y, Russell S, Mugford M. A systematic review of economic evaluation literature in Thailand: are the data good enough to be used by policy-makers? Pharmacoeconomics 2007; 25: 467-79.
5. Bergson A. A reformulation of certain aspects of welfare economics. Q J Econ 1938; 52, 310-34.
Health Technology Assessment in Developing the National List of Essential Medicines in Thailand

Visanu Thamlakitkul MD*

*Chairman of the Subcommittee for Development of the National List of Essential Medicines BE 2555-2558

The National List of Essential Medicines (NLEM) is developed in order to include the medicines that are necessary for the health needs of the Thai population. The NLEM is designed to be an “optimum list” and it is intended to be a mechanism that supports and promotes rational use of medicines for all stakeholders. It is also a reference for pharmaceutical benefits under major health insurance schemes i.e. Social Security Scheme, Civil Servant Medical Benefit Scheme and Universal Coverage Scheme.

The main principle for developing the NLEM is that each medicine to be included in the NLEM must show evidence of its efficacy/effectiveness, its benefits outweighs its risks, its efficiency (cost-effectiveness), and its budget impact is affordable. Therefore, there are three steps in the selection of the medicines to be included in the NLEM. Step 1 is to assess efficacy/effectiveness and safety of the proposed medicine. Step 2 is to assess efficiency (cost-effectiveness) of the medicine that is shown to be efficacious/effective and safe. Step 3 is to estimate the amount of needed resources if the cost-effective medicine is included in the NLEM in order to determine its affordability by the administrators of the aforementioned major health insurance schemes. All information of the three steps is submitted to the chairman of the National Drug System Development Committee who then will make a decision whether the medicine should be included in the NLEM.

The efficiency (cost-effectiveness) of the medicine that is shown to be efficacious/effective and safe has been explicitly assessed based on the guidelines of a health technology assessment (HTA) developed by a group of national experts and endorsed by the Subcommittee for Development of the NLEM since 2008.

The Guidelines for Health Technology Assessment in Thailand were revised in 2013 to include several important issues (such as the guidelines on budget impact analysis, social and ethical impact assessment) that were not discussed the first HTA guidelines. Many up-to-date methodological issues are also included in the Guidelines for Health Technology Assessment in Thailand (Second Edition).

The Guidelines for Health Technology Assessment in Thailand (Second Edition) have been approved by the Subcommittee for Development of the NLEM and they will be used for conducting health economic evaluations and budget impact analyses to consider if the medicine is cost-effective and affordable, and whether it should be finally included in the NLEM.

J Med Assoc Thai 2014; 97 (Suppl. 5): S3
Full text. e-Journal: http://www.jmatonline.com
Guidelines for Health Technology Assessment in Thailand (Second Edition)-The Development Process

Usa Chaikledkaew BSc(Pharm), MA, PhD*,**, Kankamon Kittongsiri BSc(Pharm)*,**

* 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

The first Thai-specific HTA guidelines were completed in 2008 with the aim of ensuring that all HTA data was accurate, of high quality, and relevant for making decisions pertaining to healthcare resource allocation. Based on a quality assessment of 89 economic evaluation studies in the Thai context published in international academic journals between 1982 and 2012, the analysis revealed a significant increase in quality of data sources and result reporting in studies published after the dissemination of the first Thai HTA guidelines. As the first Thai HTA guidelines were developed in 2008, a number of areas for improvement have been identified. Therefore, the objective of this chapter is to describe the development process of this second edition of HTA guidelines for Thailand which builds on the success of the first edition, while attempting to address some of the identified limitations of the first edition and reflect the changes that the health care and policy contexts have undergone in the intervening years. It is hoped that this second edition will continue to build on these successes so that policy decision making becomes increasingly evidence-based.

Keywords: Economic evaluation, Guidelines, Pharmacoeconomic

The development of health technology assessment (HTA) guidelines is a key part of any HTA research and development process. The first Thai-specific HTA guidelines were completed in 2008, with the aim of ensuring that all HTA data was accurate, of high quality, and relevant for making decisions pertaining to healthcare resource allocation. The first HTA guidelines made practical suggestions for improving the quality of health economic evaluation research and gave recommendations on key methodological issues given the particular resource and information limitations of the Thai healthcare system.

The first Thai HTA guidelines also offered guidance and resources to help researchers choose appropriate methodologies and data sources for their HTA research. By recommending a set of methodologies and data sources across HTA research, it was also hoped that there would be greater transparency, by improving the consistency and quality of research and improving research assessment, by allowing comparison against a set of pre-determined guidelines. While guidelines themselves cannot, on their own, guarantee that policy-makers will use HTA data when forming their decisions, they should help improve the quality of available HTA data and hence the extent to which policy decisions are informed by reliable scientific evidence.

The publication of the first edition of HTA guidelines has been recognized as a significant step forward in improving HTA research and development (and thus policy-making). They have received widespread acceptance in the scientific and policy making communities in Thailand, and have been approved by the Subcommittee for Development of the National List of Essential Drugs and the Subcommittee for Development of the Health Benefit Package and Service Delivery of the National Health Security Office. However, it is important to note that guidelines such as these should always be regarded as dynamic tools that will require refinement over time.

Correspondence to:
Chaikledkaew U, Social and Administrative Pharmacy Excellence Research (SAPER) Unit, Department of Pharmacy, Faculty of Pharmacy, Mahidol University, 447 Sri-Ayudthaya Road, Payathai, Ratchathewi, Bangkok 10400, Thailand. Phone: 0-2644-8678 ext. 5317, Fax: 0-2644-8694 E-mail: usa.chi@mahidol.ac.th
and should be adapted according to the changing nature of Thailand’s healthcare context. Since the publication of the first edition, a number of areas for improvement have been identified, and the need for a second edition that reflected the changing context was recognized.

The first Thai HTA guidelines focused exclusively on making HTA recommendations from an economic standpoint; no consideration was given to other issues that have been deemed relevant for HTA, such as budgetary, social, and ethical impacts. In addition, it was widely agreed that the first Thai HTA guidelines, particularly the theoretical justification, were somewhat complicated and hard to understand. In developing the second set of guidelines, it was agreed that greater attention should be given to HTA application rather than theory, to ensure that the content was relevant and easily comprehensible.

In addition, as the first Thai HTA guidelines were developed in 2008, a number of elements were identified as out of date. For example, they included no reference to a standard cost list (a reference tool for cost assessment) and no information on the current cost-effectiveness threshold in Thailand. The first set of guidelines also relied on the EQ-5D questionnaire as an assessment of quality of life, which—while useful—has been replaced in many areas by the recently-developed, more refined assessment tool, the EQ-5D-3L questionnaire. The guidelines also make no mention of indirect comparison meta-analysis, a tool that is widely used for the indirect comparison of the clinical outcomes from a randomized controlled trial, nor do they make reference to the economic evaluation of specific conditions that require specific methodologies to simulate the progression of the disease (such as certain infectious diseases, which require dynamic models). They also give no guidance on interventions such as screening and diagnosis or medical devices.

The development of this second edition of HTA guidelines for Thailand builds on the success of the first edition, while attempting to address some of the identified limitations of the first edition and reflect the changes that the health care and policy contexts have undergone in the intervening years.

**Developing the second edition**

The development of the second set of HTA guidelines began in June 2012, when researchers at the Health Intervention and Technology Assessment Program (HITAP) undertook a systematic review of health technology assessment research in the Thai context. The researchers focused particularly on economic evaluations and compared those published before the introduction of the first Thai HTA guidelines (1982 to December 2008) with those published afterwards (January 2009 to September 2012). To investigate how effective the HTA guidelines were, the researchers examined the extent to which the methodology of studies published after the introduction of the guidelines was consistent with the recommendations given. As well as comparing the studies with the recommendations, quality was also assessed using the quality assessment framework developed by Teerawattananon et al(1), and by assessing the quality of reporting (using criteria developed by Drummond et al(2,3)) and quality of data sources (using criteria developed by Cooper et al(4)) because the results from these criterion could reflect the quality of the studies. Fig. 1 shows the process through which the second edition of Thai HTA guidelines was developed.

In July 2012, HITAP, the body that serves as the coordinator of Thai HTA guidelines, arranged a consultation meeting for experts and stakeholders, including the Subcommittee for Development of the National List of Essential Drugs, the Subcommittee for Development of Health Benefits Package and Service Delivery of the National Health Security Office, graduate students, professors, researchers, academics, and other public and private stakeholders. During the meeting,
participants discussed what, aside from economic concerns, should be taken into consideration when developing the second set of HTA guidelines. They also explored the findings of several HTA studies conducted by HITAP and discussed what implications these studies might have for the development of the second edition. The benefits and limitations of the first set of guidelines were discussed, and a number of issues were identified, including the need for content modernization and the fact that there was, at present, inadequate information on how to apply the guidelines. The second HTA guideline development working group then revised the guidelines in light of all of the comments and suggestions that had been made. In November, 2012, a second expert and stakeholder consultation was held to consider the appropriateness of the draft version of the second set of guidelines. The comments and suggestions from this meeting were then incorporated into the draft to produce a final version.

**Conceptual framework**

Health technology assessments are policy research tools that investigate the long- and short-term effects of health technologies in a systematic way from a multidisciplinary standpoint. It is used to capture the whole range of effects—direct and indirect, intentional and unintentional, any effect resulting from the development, diffusion, and application of the technology etc.—so that decisions can be made about the value of a given technology. Following the initial consultation with experts and stakeholders that was held in July 2012, it was agreed that the second Thai HTA guidelines should be developed with the aim of increasing the quality and standard of HTA in Thailand, rather than as a manual for performing HTA research. In addition, the second Thai HTA guidelines should build on the first guidelines by looking at health technology assessment from a budgetary, social, and ethical standpoint, as well as an economic standpoint.

The second edition of the Thai HTA guidelines gives greater guidance on how the guidelines should be applied, including recommendations on which data sources and tools are most appropriate based on explanatory examples from HTA research. The guideline document is concise, consisting of an introduction, an outline of concepts and principles, and a summary of recommendations. HTA theory is referred to only briefly in this second edition, since this was covered in detail in the first edition; instead, the focus of this latest set of guidelines was the improvement of research quality by increasing the availability of high quality HTA data in the hope that it will play an increasing role in the formation of policy. An overview of the content of the 2012 publication—“Guidelines for Health Technology Assessment in Thailand—Second Edition”—the final document that emerged from the process, is shown in Fig. 2 below.

**Quality assessment of HTA research**

A quality assessment was conducted in January 2013 of 89 economic evaluation studies in the Thai context published in international academic journals between 1982 and 2012. The studies were divided into two groups—those that were published before the dissemination of the first Thai HTA guidelines (January 1982 to December 2008) and those that were published after (January 2009 to September 2012). Quality assessment data on studies conducted between 1982 and 2005 was taken from Teerawattananon et al’s 2007 review. Their systematic review of literature relating to Thailand revealed a number of methodological flaws with previous HTA publications. The review highlighted that serious attention needed to be given to the quality of reporting and the use of information in the analyses. In addition, it demonstrated...
significant variation in the methods that were used, making the comparison of results between studies very difficult. One way in which Teerawattananon et al suggested tackling these challenges was through the establishment of standard guidelines for conducting HTA.

Following the implementation of the first Thai HTA guidelines, a comparative analysis was conducted to examine whether research quality increased after the publication of the guidelines. This was done by assessing the extent to which important issues recommended in the first Thai HTA guidelines were adopted by subsequent studies, and by analysing quality of reporting according to criteria developed by Drummond et al\(^\text{[2,3]}\) and quality of data sources according to criteria developed by Cooper et al\(^\text{[4]}\). Key factors that were examined in the analysis included whether the study had adopted a clearly-defined perspective and whether the present study compared two or more interventions. The use of incremental cost-effectiveness ratios (ICERs), uncertainty analyses, and discounting methods (where the study period was longer than one year) were also assessed for relevance, and the implications of any funding support were taken into account. Using the scale developed by Cooper et al, all data sources that were used were ranked from 1-9 according to their reliability (where 1 is most reliable).

**Assessing reporting quality**

Table 1 shows the results of the comparative analysis that was conducted on the quality of research reporting in economic evaluation studies published in international journals before and after the dissemination of the first Thai HTA guidelines. The criteria that was used was taken from Drummond et al\(^\text{[2,3]}\). The analysis revealed a significant increase in quality in studies published after the dissemination of the first Thai HTA guidelines.

<table>
<thead>
<tr>
<th>Reporting issues</th>
<th>Dissemination period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear identification of perspective used in the study</td>
<td>28/52 (54)</td>
</tr>
<tr>
<td>Identification of compared interventions</td>
<td>44/50 (88)</td>
</tr>
<tr>
<td>Discounting method used for future cost and outcomes, if study period is longer than one year</td>
<td>9/23 (39)</td>
</tr>
<tr>
<td>Result presentation as incremental cost-effectiveness ratio (ICER)</td>
<td>22/49 (45)</td>
</tr>
<tr>
<td>Performing uncertainty analysis</td>
<td>21/52 (40)</td>
</tr>
<tr>
<td>Disclosure of funding support</td>
<td>35/52 (67)</td>
</tr>
<tr>
<td></td>
<td>33/37 (89)</td>
</tr>
<tr>
<td></td>
<td>37/37 (100)</td>
</tr>
<tr>
<td></td>
<td>22/24 (92)</td>
</tr>
<tr>
<td></td>
<td>28/33 (85)</td>
</tr>
<tr>
<td></td>
<td>34/37 (92)</td>
</tr>
<tr>
<td></td>
<td>27/37 (73)</td>
</tr>
</tbody>
</table>

**Table 1.** Quality of research reporting before and after dissemination of Thai HTA Guidelines

**Assessing data source quality**

Table 2 shows the results of the comparative analysis that was conducted on the quality of data sources economic evaluation studies before and after the dissemination of the first Thai HTA guidelines. The criteria that was used to assess the quality was taken from Cooper et al\(^\text{[4]}\). The quality of all data sources was ranked from 1 (best) to 6 (worst) and 9 (not stated). Important data sources that were evaluated included clinical effect size, baseline clinical data, adverse events and complications, resource use, cost and utility. The analysis revealed an increase in quality in a number of data sources following the dissemination of first Thai HTA guidelines, including clinical effect size, adverse events and complications, baseline clinical data (for more details on the ranking of data sources, including those concerned with clinical effect size, and adverse events and complications, see Measurement of Clinical Effects on the first guidelines\(^\text{[5]}\)).

However, the quality of some data sources, namely resource use and cost, was found to have decreased after the dissemination of the guidelines. This is probably due to the fact that most researchers were compelled to collect their own resource utilization and cost data due to a lack of published research at that time—leaving significant room for inconsistency across studies. As the number of studies on resource use and cost has increased, it is likely that quality in these fields will increase too. The quality of utility data sources was found to have increased following the dissemination of the first Thai HTA guidelines (for more details, see Measurement of Utility on the first guidelines\(^\text{[5]}\)).

The preliminary data suggests that, following the dissemination of the first Thai HTA guidelines, the quality of the reporting and data sources used in economic evaluation studies in Thailand increased.
significantly, and other areas of research showed improvement. These improvements suggest that HTA guidelines can play a valuable role in improving research quality, and it is hoped that this second edition will continue to build on these successes so that policy decision making becomes increasingly evidence-based.

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 organisations 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


2. Teerawattananon Y, Russell S, Mugford M. A systematic review of economic evaluation literature in Thailand: are the data good enough to be used by policy-makers? Pharmacoeconomics 2007; 25: 467-79.


กระบวนการพัฒนา "มือการประเมินแพทย์ไทย" ด้านสุขภาพส่วนบุคคล ประเทศไทย ฉบับที่ 2

ลูกา อาเก็เดิมเกะ, ทีมคณะ ศิลปสหวิทย์

คู่มือการประเมินแพทย์ไทยด้านสุขภาพส่วนบุคคล ฉบับที่ 1 ดำเนินการแล้วเสร็จ และได้รับการตีพิมพ์เผยแพร่ต่อเนื่อง พ.ศ. 2551 โดยมีวัตถุประสงค์เพื่อเป็นแนวทางสำหรับผู้ผลิตและผู้ใช้แรงบันดาลเพื่อใช้ในการตรวจสอบความถูกต้องและความถูกต้องของงานวิจัยที่ใช้สำหรับการจัดสรรและสร้างคุณภาพของงานวิจัยโดยเป็นที่ยอมรับ จากการประเมินคุณภาพของกระบวนการพัฒนา "มือการประเมินแพทย์ไทย" สำหรับการณ์การประเมินคุณภาพคู่มือสุขภาพที่เกี่ยวข้องกับประเทศไทยที่พิมพ์ในวารสารวิชาการทางการศึกษาแพทย์ พ.ศ. 2525 ถึง พ.ศ. 2555 จำนวนทั้งสิ้น 89 ฉบับ พบว่ามีปัญหาเกี่ยวกับระดับการเผยแพร่ คู่มือฯ ฉบับที่ 1 งานวิจัยด้านการประเมินความคุ้มค่าด้านสุขภาพคู่มือสุขภาพไทยในระดับการแพทย์สุขภาพที่ ฉบับที่ 1 มีคุณภาพเพียงพอต่อการมีคุณค่าที่ดีในลานการทำงานการพยาบาลและสุขภาพของมนุษย์ อย่างไรก็ตามเนื่องจากคู่มือฯ ฉบับที่ 1 ได้รับการพัฒนาต่อเนื่อง พ.ศ. 2551 เนื่องจากหลายส่วนของหน้าต่างรายเปรียบปรุง ดังนั้น ต้องประสบความทุกข์ที่ถึงมีช่องว่างบางอย่างรวดเร็วทางการพัฒนา "มือการประเมินแพทย์ไทย" ฉบับที่ 2 ซึ่งเปรียบปรุงจนทั่วคู่มือฯ ฉบับที่ 1 โดยพยายามแก้ไขจุดอ่อนที่มีอยู่ในคู่มือฯ ฉบับที่ 1 และสะดวกรับทราบระบบสุขภาพและนโยบายด้านสุขภาพที่เปลี่ยนแปลงไป โดยหวังเป็นอย่างยิ่งว่าคู่มือฯ ฉบับที่ 2 นี้จะช่วยผลักส่งให้มีการกระจายความรู้ภายในระบบทางการแพทย์และระบบสุขภาพที่มีอยู่ในประเทศไทย.
Defining the Scope of Health Technology Assessment and Types of Health Economic Evaluation

Pimwara Tanvejsilp MS*,
Surachat Ngorsuraches PhD*

*Faculty of Pharmaceutical Sciences, Prince of Songkla University, Hat Yai, Songkhla, Thailand

Health Technology Assessment (HTA) is a process that uses principles from across various disciplines, including medicine, sociology, economics, and ethics, to evaluate health technologies. Policy makers can use HTA as a tool to assess health technologies in a systematic, unbiased, transparent, and robust manner in order to make informed and evidence-based decisions. Generally, researchers begin an HTA by defining the overall scope of the assessment, after which they choose an appropriate type of health economic evaluation, one of the most important elements of HTA. The objective of this article is to provide recommendations about the scope of HTA as well as guidance on the kinds of health economic evaluation that are appropriate in the context of the new Thai HTA guidelines. A well-defined research question that addresses five major components−target population, technology or intervention, comparator, outcome of interest, and perspective−is an essential part of any HTA.

Keywords: Health technology assessment, Scope of health technology assessment, Health economic evaluation

Health Technology Assessment (HTA) is a process that uses principles from across various disciplines, including medicine, sociology, economics, and ethics, to evaluate health technologies. Policy makers can use HTA to assess health technologies in a systematic, unbiased, transparent, and robust manner, so that they can make informed and evidence-based decisions. In recent years, four main areas have emerged among those interested in HTA−policy analysis, evidence-based medicine, health economic evaluation, and social and humanistic impact assessment(1). However, while these areas have been widely accepted, it is important also to acknowledge that not every HTA will necessarily include consideration of all four.

HTA is usually initiated at the beginning of a policy process so that the findings can be used to inform at the various steps of the process e.g. agenda setting, policy formulation, decision, and evaluation. In this way, policy makers are able to use HTA to help them make evidence-based decisions and, eventually, to ensure that cost-effective technologies are given priority over those which have doubtful value for the health system. In the last twenty years, the importance of using evidence-based decision-making when assessing the efficacy or effectiveness of health technologies has been increasingly recognised. In general, the evidence that is used in HTAs is experimental, quasi-experimental, or comes from observatory studies. Recently, comparative effectiveness data from head-to-head clinical trials or comparisons using both clinical and management data, has been used to inform clinical guidelines and health insurance benefit coverage guidelines. However, these kinds of evidence usually do not include information on the cost-effectiveness of the technologies under consideration. This is where health economic evaluations can help. Health economic evaluations assess the trade-off between the resources necessary to adopt a certain technology and the benefits of that technology once it is adopted. In essence, an HTA investigates the extent to which the adoption of a specific technology is cost effective. While economic evaluations are only one part of HTA, they have historically been the aspect that has been focused on the most. However, it is important to recognize that HTAs involve more than just an assessment of the benefit of a technology for an associated resource use. HTAs should include not only an assessment of the medical evidence behind and an economic evaluation
of the technology in question, but also the impact of that technology on organisations, society, ethics, and humanity\(^1\).\(^2\)

All HTAs should begin by defining the overall scope of the assessment and the research framework that will be utilised, based on general HTA concepts and principles. For instance, a clear HTA research question should be defined, based on the policy context and a review of the existing literature. It is also important that researchers clarify the perspective that they will be adopting in their research. Since health economic evaluations are a major part of any HTA, this article briefly defines the four main types—Cost-Minimisation Analysis (CMA), Cost-Benefit Analysis (CBA), Cost-Effectiveness Analysis (CEA), and Cost-Utility Analysis (CUA). The basic concepts of the incremental cost-effectiveness ratio (ICER) and the cost-effectiveness plane are also presented.

**Concepts and principles of HTA**

This section intends to provide an overview of the general concepts and principles of HTA. It gives guidance on the scope of HTA, provides details on the kinds of research framework that are used and the principles that lie behind technology selection. It also provides a summary of the types of health economic evaluations that are used in HTA and some guidance on how to interpret HTA results.

**Defining the scope and research framework**

**Health technology selection for HTA**

In any healthcare context, it is usually not possible to assess all health technologies due to resource limitations. Indeed, only some health technologies will be appropriate for HTA. Generally, the selection of a technology for assessment depends on the priorities of the organisation undertaking the assessment. However, there are two important principles that should always inform the decision when selecting a health technology for assessment to be rigorous, evidence-based, and transparent\(^3\)-\(^5\).

Selection criteria: Although most countries do not have a well-defined set of criteria to help guide health technology selection for HTA, there is widespread consensus that any technology that has been shown to offer significant health benefits or impact significantly on health-related policy e.g. in terms of equity or patient access, is worthy of assessment. Moreover, most countries agree that consideration should also be given to those technologies that treat diseases that are particularly virulent and could create major financial losses to either individuals or society. Finally, there is also agreement that technologies that have a particularly wide application and those that are in particularly high public demand should also be considered for HTA.

Selection process: After selection criteria are identified, different countries have their own different selection processes, e.g. priority setting by scoring system, expert opinions, etc. However, transparency and participation are common principles of health technology selection for HTA.

**Defining the research question**

In essence, HTA is a policy research tool\(^6\). The first step in undertaking an HTA therefore, is to decide which research question is suitable, given the policy question under discussion. Usually, well-defined research questions emerge from a collaborative discussion between stakeholders and researchers. A good research question will always include five components\(^7\),\(^8\).

- **Target population**: This defines the population of interest by defining various characteristics. The characteristics that are specified should always include age, gender, and disease risk, as a minimum.
- **Technology or intervention**: This will specify the type of technology under discussion. Details on how innovative the technology is, will be given here and any existing alternatives to the technology should be outlined.
- **Comparator**: Similar details indicated for technology or intervention, which is chosen to be compared with the technology or intervention of interest, should be provided.
- **Outcome of interest**: The expected outcomes should be clearly indicated.
- **Perspective**: The type of perspective that is to be used should be indicated.

In addition, a well-defined research question should include an outline of the overall research context within which the assessment is taking place i.e. by defining who is calling for the HTA, identifying the primary stakeholders (policy makers, health care providers, patients, etc.), and clarifying how this HTA is expected to be beneficial.

**Health technology background review**

Reviewing the existing literature and information on the condition and treatment under investigation is an important part of any HTA.
Researchers should present a clear contextual description of the condition, including an outline of its consequences, prognosis and progression, clinical characteristics, period, and treatment alternatives. Researchers should always be aware that their readers may not be clinical experts. Researchers should also provide an overview of any other background information that will be relevant to the HTA, including summarising any safety issues that have emerged, as well as providing an overview of the existing data on the health technology’s efficacy and effectiveness. The details of the outcomes and its measurement should also be clearly described.

HTAs should always include a health technology background review—not only for technologies such as pharmaceuticals, but also for medical devices, community interventions, and other medical procedures. The health technology background review should examine both the technology itself, as well as relevant factors that affect its implementation e.g. facilities, procedures of care, distributions of technology, indications, price, legal issues, manufacturers, market shares, etc.

**Target population**

In the target population section of the HTA, researchers should begin by providing an overview of the disease’s epidemiology, including prevalence and incidence. This helps to give a general idea of the extent to which the disease condition affects the general population. The target population for the HTA should then be clearly defined, including their age, gender, socioeconomic status, and risk factors. This means that the results of the study can be linked clearly to the appropriate population, rather than the general population, which is important in determining the extent of cost-effectiveness (i.e. whether a treatment is cost-effective for the general population, or only to the target population). These kinds of subgroup analysis also help to reduce bias.

**Selection of comparators**

The selection of appropriate comparators is also an important step when undertaking an HTA, especially in terms of the health economic evaluations. Having appropriate comparators helps guide how the HTA results should inform practice and prevents bias, by ensuring that relevant alternative technologies are also assessed for cost-effectiveness. Any technology used in current practice, which is agreed by the relevant stakeholders, can be a comparator. While comparators are often technologies that are direct alternatives to the technology under consideration, all possible alternatives should be considered and assessed until the most appropriate comparators are found. Indeed, although the technologies that are most widely used in the treatment of the condition under examination are often used as comparators, they may not always be the best choice, particularly if the widest-used technology is not the most effective technology. The technology that has been shown, to date, to be most effective (especially those that are recommended by standard practice guidelines and those recommended by experts), should be used as a comparator, where possible. It is also important to note that the watchful waiting option (WAW), rather than implementing an alternative is often the most appropriate comparator. There is no restriction on the number of comparators that can be included in an HTA; the number will depend on various limitations, including data and time availability.

**Perspective**

Defining the perspective that the HTA will adopt is an important part of the HTA process, since this will define, to a certain extent, the different costs and outcomes of the assessment. For instance, the broadest societal perspective will take into account the direct and indirect costs of both the health care system and the patients. In contrast, the narrower perspective will focus only on the costs of one body—e.g. a health ministry or a hospital, and will exclude all patient costs.

Even though there is no such thing as a “best perspective”, researchers should define their perspective based on the objectives of the HTA user. In general, however, broader perspectives tend to give more informative results, which maybe more useful for those involved in policy decision-making.

**Health economic evaluation methods**

**Types of health economic evaluation**

Limited health care budgets mean that decisions about the amount of resources that can be allocated to health technologies have to be made. In policy decision-making, health economic evaluation is an important tool that is used to compare various treatment alternatives to ensure that policy-makers are able to allocate health care resources in the most efficient way.

A health economic evaluation is a comparative analysis of a number of treatment options. It focuses
on two factors—costs and outcomes\(^{(8)}\), Drummond et al have identified four types of health economic evaluation: Cost-Minimisation Analysis (CMA), Cost-Benefit Analysis (CBA), Cost-Effectiveness Analysis (CEA), and Cost-Utility Analysis (CUA). Table 1 summarises the cost and outcome definitions and measurements of these four types.

**Cost-minimisation analysis (CMA)**

CMA is an analysis that compares two or more alternatives that have demonstrated equivalence. It is used to show the alternative that has the lowest costs\(^{(7)}\). While useful, there have been some criticisms levelled at the approach. For instance, Briggs et al identified relatively high levels of uncertainty in the process of cost and outcome estimations\(^{(11)}\). This means that the only technologies that can be meaningfully compared using CMA are those that are really equivalent. Given this, Drummond et al have recently suggested that CMA be used only as a cost analysis tool, rather than a full economic evaluation\(^{(8)}\).

**Cost-benefit analysis (CBA)**

CBA is an analysis that compares technologies by converting outcomes to monetary units and then comparing the costs and resources used (e.g. by calculating amounts such as the costs saved as a result of using the technology of interest, and the costs that were avoided as a result of early diagnosis). In some CBA cases, the willingness-to-pay (WTP) of patients is included, so that the analysis is informed by the maximum amount of money that patients would be happy to pay to treat or prevent the condition\(^{(8,10)}\). Since both costs and outcomes are measured in monetary units, any technology where the costs are found to be lower than the value of outcome is cost-effective. This method allows easy comparison between two or more technologies, as all costs and outcomes are converted into the same monetary unit. However, the conversion of health outcomes to the monetary unit is challenging and controversial. It limits the use of CBA in health economic evaluation.

**Cost-effectiveness analysis (CEA)**

CEA is widely used in health economic evaluations to ascertain the most efficient way to allocate resources. This approach measures costs in monetary units and outcomes in efficacy or effectiveness units\(^{(10)}\), (e.g. cases of correct diagnosis, deaths, life-years gained, etc.) and then compares the value of a given treatment, given these measurements. Outcomes can be quite specific (e.g. diastolic blood pressure reduction, reduction in the number of pain points, etc.), which allows for quite detailed tailored analysis. Intermediate outcomes (e.g. detected risk factors) and final outcomes (e.g. cases prevented) are also sometimes included. However, the use of intermediate outcomes in CEAs is not recommended, since this may result in biased interpretation of study results due to its limited use in predicting final outcomes. A major limitation of CEA is that it cannot be used to compare health technologies with different outcomes.

**Cost-utility analysis (CUA)**

CUA is used when the outcome of interest is health-related quality of life. The most widely used outcome measurement in CUA is the quality-adjusted life year (QALY), which captures both quantity (life

| Table 1. Types of Health Economic Evaluation\(^{(8,10)}\) |
|----------------|-----------------|-----------------|-----------------|
| Type of analysis | Cost measurement | Outcome characteristics | Outcome measurement |
| Cost-minimisation analysis | Monetary unit | Every outcome is equivalent. | Assuming equivalent outcomes, measurement is not necessary. |
| Cost-benefit analysis | Monetary unit | Every outcome may not be the same because it must be converted to monetary unit. | Monetary unit |
| Cost-effectiveness analysis | Monetary unit | Every outcome must be in the same unit. | Efficacy e.g. life years gained, number of correctly diagnosed patients, reduced blood pressure, etc. |
| Cost-utility analysis | Monetary unit | Every outcome may not be the same because it must be converted to utility. | Quality-adjusted life years (QALYs) |
year) and quality (quality of life as a utility measurement) of outcomes. In general, utility score varies from zero (death) to one (full health).

To undertake a QALY calculation, the utility score of a health state would be multiplied by the time the patient spent in that state. For instance, if a patient has full health for the first six months (utility score 1), followed by six months where they suffer from condition Z (utility score 0.5), this patient would have gained a 0.75 QALY.

Tarride et al found that the methodology that is used to determine utility scores or health preferences is an important part of CUA(10). Generally, there are two measurement methods—direct and indirect. Standard gamble (SG) and time-trade off (TTO) are examples of the direct method, but they use a lot of resources and can be time-consuming to undertake. Euro Qol (EQ-5D) and Health Utility Index (HUI) are examples of the indirect method. QALY is not only the measurement outcome that is used in CUAs. Another commonly used outcome is the disability-adjusted life year (DALY)(8), which was developed by the World Bank and Harvard University and, subsequently adopted by the World Health Organisation (WHO) in 1996.

Incremental cost-effectiveness ratio (ICER)

While choosing the appropriate type of health economic evaluation is an important part of HTA, the way that results are interpreted is also important. It would, of course, be easy to make decisions regarding superior treatment if all new technologies were superior to older ones in terms of better health outcomes and cheaper costs. However, the reality is that, while most new technologies tend to give better health outcomes, they also come with higher costs. Therefore, policymakers need to make a trade-off between costs and health outcomes. Tarride et al has suggested that the goal of any health economic evaluation is to determine how much more it is appropriate to pay for a unit of a health outcome, when a new technology replaces an existing technology(10). The incremental cost-effectiveness ratio (ICER) is an effective way to measure this; it is based on a calculation that determines the difference between the costs of old and new technologies, divided by the difference between their outcomes. The formula for determining an ICER value is shown below:

\[
    \text{ICER} = \frac{(\text{Cost}_A - \text{Cost}_B)}{(\text{Outcome}_A - \text{Outcome}_B)},
\]

where A represents higher cost and better outcome technology, and B represents lower cost and poorer outcome technology. For instance, if a new technology and an old technology cost 200,000 Baht and 140,000 Baht, respectively, and the new technology results in a gain of 1 QALY while the old technology results in only a 0.5 QALY gain, then the ICER value will be 120,000 Baht \((200,000-140,000)/(1-0.5))\) per QALY. This means that, in this case, the new technology will cost 120,000 Baht per QALY more than the old technology for equivalent outcomes. Another kind of measurement that is sometimes used to calculate results in an HTA is the average cost-effectiveness ratio (ACER). It is often confused with ICER. The main differences between the two types of equation are given in Table 2 below.

In health economic evaluations, ICER is used to indicate whether a technology is cost-effective or not. The ACER, on the other hand, is not appropriate for making these kinds of decisions, since the ratio compares the technology of interest with a do-nothing alternative, which has no cost and no health outcome. In practice, WAW is rarely a good comparator, since it not only is it rarely used in the practical healthcare setting, but its usage can also be ethically problematic. In addition, WAW often generates more costs, since the lack of treatment can result in more severe symptoms and more costly in the future.

Another tool that is useful in conducting health economic evaluations is the cost-effectiveness plane(10). This is a graphical plane between the incremental cost and incremental outcome. If we apply this to the above-mentioned example, then the line would be plotted as shown in Fig. 1. The horizontal axis represents the incremental outcome while the vertical axis represents the incremental cost between new and old technologies. The cost-effectiveness plane has four quadrants—A, B, C, and D. Quadrant A is the area of the ICER where the old technology is the dominant alternative, while quadrant D is the area where

<table>
<thead>
<tr>
<th>Cost(Baht)</th>
<th>QALY</th>
<th>ACER</th>
<th>ICER(Baht/QALY)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>200,000</td>
<td>1</td>
<td>200,000/1</td>
</tr>
<tr>
<td>B</td>
<td>140,000</td>
<td>0.5</td>
<td>140,000/0.5</td>
</tr>
</tbody>
</table>

Table 2. ACER and ICER Examples
the new technology is the dominant alternative. Quadrant B and C are areas that show the trade-off between costs or resources used and outcomes obtained since they are the areas where the new technology has higher costs and better outcomes or lower costs and poorer outcomes. The cost-effectiveness plane can help policy-makers quickly understand the various benefits and drawbacks of a new technology and its comparator.

**Guidelines for health technology assessment in Thailand (second edition): Recommendations for defining the scope of HTA**

Given the concepts and principles of HTA and health economic evaluations that have been presented here in, several recommendations are made for researchers in terms of defining the scope of HTA and the types of health economic evaluation that are appropriate for the Thai context.

1. All HTAs should include a well-defined research question that incorporates five major components—target population, technology or intervention, comparator, outcome of interest, and perspective.

2. All HTAs should include a comprehensive background information review that summarises all relevant information on the disease and health technology in question.

3. The target population in every HTA should be clearly described.

4. All comparators should be technologies that are used in current practice, and which have been discussed or approved by stakeholders. They should also be technologies that may be replaced by the technology of interest. Regardless of the type of comparator that is used, clear reasons and details must be provided.

5. The societal perspective should be adopted.

6. A Cost-utility analysis (CUA) is recommended to be the method of choice. If data or resources are limited, CEA can be used. However, the use of intermediate outcomes is not recommended.

**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 organisations 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**


4. Goodman CS. HTA 101 Introduction to health care


Measurement of Costs for Health Economic Evaluation

Arthorn Riewpaiboon BSc(Pharm), MSc, PhD*

* Social and Administrative Pharmacy Excellence Research (SAPER) Unit, Department of Pharmacy, Faculty of Pharmacy, Mahidol University, Bangkok, Thailand

The provision of guidelines on cost measurement for health economic evaluations enable research to be more standardized and hence more comparable, which offers clear benefits for policy formulation and health management. The guidelines herein focus on three aspects-the cost of health intervention/health care programs, the cost of illness/health risks, and use of costs in health economic evaluation. For each aspect, the main concepts and methods are outlined, and recommendations for the Thai context are presented. There is particular focus on how to calculate various costs according to different evaluation methods and perspectives, how to evaluate source of cost data, how to make value adjustments and how to present cost measurement findings.

Keywords: Cost, Cost measure, Methods, Guidelines, Economic evaluation, Thailand

‘Costs’ refer to the monetary value of resources that are used for the production of products or services, based on a concept of opportunity cost. In health economics, measuring or calculating costs, conducting a cost analysis, or undertaking costing research are processes that help researchers know or estimate the total cost of a service or technology, by calculating the cost components and costs per unit of outputs. Some costs analyses also include an investigation of the factors that affect those costs. By calculating the costs of technologies, researchers, health administrators, and policy-makers are better able to prioritize health problems as well as implementing more evidence-based financial and budgeting management, efficiency management, and health economic evaluations. This article is concerned with how cost calculations affect health economic evaluations. The author hopes that this overview will provide researchers with a standard model and process that will help generate more uniform and comparable research in the future and ultimately benefit policy formulation and health management.

Health economic evaluations take into account costs involved with the provision of medical services/health care programs, any costs associated with the illness and/or health risks (or economic outcome), and how best to apply costing data. Health economic evaluations are a key tool in the health system management that help improve efficiency and equity(1) as well as informing decisions about equality and ethics. For instance, when calculating time costs, the same rate is used for all individuals, regardless of gender, age, and income to avoid ethical violation.

Measuring the costs of medical services and healthcare programs

To begin, the author examines the process for estimating the cost of the two main types of health interventions-medical services and healthcare programs. The author will consider two key approaches for measuring the costs involved with medical services-direct unit cost analysis and standard unit cost.

Unit cost analysis of medical services

There are two main methods for conducting a unit cost analysis-standard costing(2-4) and activity-based costing(5-7). Standard costing is more well-known and is widely recognized as being more appropriate for the Thai context. The standard costing method is composed of six steps: study design and planning, cost centre classification, direct cost determination, indirect cost determination, full cost determination and unit cost calculation. The author shall examine each of...
these in turn.

**Step one: Study design and planning**

The study design and planning step involves a number of different elements: The objectives must be defined, the cost objects (or cost products) and their components must be identified, the perspective and time horizon that will be adopted must be agreed, and the level of organization involved must be taken into account.

**Step two: Cost centre classification**

A cost centre is a unit that produces output and has a record of resource consumption. Cost centres can be categorized in a number of ways. However, in the process of unit cost calculation, all cost centres are classified into two groups:

Transient cost centres are those that provide supporting services to other cost centres. Direct costs of these cost centres are allocated to other cost centres they support.

Absorbing cost centres are cost centres that produce cost objects. These cost centres receive the costs that are allocated to them by transient cost centres. This step involves examination of the structure of the hospital so that it can be broken down into cost centre components (usually hospital departments). The smaller the cost centres and the fewer service types each individual cost centre produces, the more accurate the analysis.

**Step three: Direct cost determination**

In step three, the total direct costs are calculated. The calculation includes all the cost components that were identified in step one. The calculation includes all material costs (drug, medical materials, office materials, utilities, maintenance, hiring, and outsource); all labour costs (salary, over-time, welfare and other compensation); all capital costs (capital asset costs—for instance, building, furniture, equipment, vehicles etc). The calculation takes into account the asset’s use in the calculation period as well as the opportunity costs for any remaining payment that will be made for remaining future years\(^9\). The calculation period or useful year is defined according to recommendations from the Ministry of Finance. In practice, this is already defined in the capital asset registration.

**Step four: Indirect cost determination**

Indirect costs are all costs from transient cost centres that have been allocated to absorbing cost centres. Costs are allocated according to several methods, the most accurate of which is the simultaneous equation method\(^8\). The allocation process begins by identifying a set of allocation criteria, as determined by the output of the transient cost centres. Where transient cost centres produce more than one output, the output onto which the criteria should be applied will be the output that is responsible for the majority of cost consumption. The quantity of this output (allocation criterion) used by all cost centres is measured to calculate allocation proportion. Finally, this proportion is used in the allocation method selected.

**Step five: Full cost determination**

The full cost of the absorbing cost centre is calculated by adding the indirect costs to the direct costs, as incurred by the absorbing cost centres.

**Step six: Unit cost calculation**

In the case where the absorbing cost centre produces only one output (a cost object) or a number of homogeneous outputs (for instance, out-patient service), average unit costs are used. For multi-product cost centres, a number of methods are available, the most accurate of which is the micro-costing method, since this is based on actual resource use\(^9,10\). This method first determines the direct cost of each service (the amount of countable resources that are used in the provision of the service). Following this, a calculation is made of the indirect cost of services (the full cost of each department subtracted by the sum of the total direct costs of all services), which is then allocated to each service using either the average method or by calculating the proportion of the direct cost of each service. The second method is the ratio of costs to charges method\(^8,11\). The ratio of cost to charges is computed based on historical records. It is used to estimate the cost of each service from the relevant charge information obtained from patient bills. This calculation can be fairly simple. For instance, in any given period, if total charges are 100,000 baht and total costs are 75,000 baht, the cost to charge ratio (CCR) will be 0.75. The ratio will then be used for determining the costs for services. For example, each service charge would be multiplied by 0.75, resulting in a unit cost of each service. The third method is the relative value unit method\(^10,11\), which uses the standard relative value unit (RVU) of each output in the calculations. The total number of all RVUs for all outputs is calculated (standard RVU of output x total number of
such output) and the cost per RVU is then calculated by dividing the total hospital cost by the total hospital RVUs. Finally, the cost per RVU is multiplied by the number of RVUs of each output.

**Standard cost list**

To increase the usefulness of health economic evaluations for national policy formulation, the evaluation results should reflect nationwide trends and be comparable with findings from other studies. To achieve this, unit costs used in cost measurement should be standardized, and to this end, a list of standard unit costs has been developed. The list is composed of unit costs for medical services in district hospitals and provincial/regional hospitals, costs of transportation and meals, and includes the time cost of outpatient visits and treatment at health centres, district hospitals, and provincial/regional hospitals. For further detail, see the relevant article in this journal.

**Cost analysis of healthcare programs**

Healthcare programs are sets of activities conducted to improve healthcare in a given population. These often involve multiple agencies and bodies who work together. For instance, in Thailand the universal vaccine program is undertaken collaboratively by the National Health Security Office, district hospitals, and local health centres; the influenza mitigation and control program is implemented by central, regional, provincial and district organizations in both the health and non-health sectors. These programs can be temporary or permanent and can involve staff from several different organizations. They usually have a specific resource and budget allocation that can come from the program itself or the organization responsible for implementing the program. For instance, one pilot vaccination program receives funding from both the sponsor organization and the ministry of public health. There are seven stages in costing a healthcare program: 1) Designing the study and planning, which takes into account defining the study objective, the perspective, the time horizon, and the types of cost that will be included (whether full or incremental costs will be used), 2) Defining the activities and organization that are involved in the healthcare program, 3) Defining the cost components. Full cost refers to the total cost of all resources used. Incremental costs refer only to the extra costs incurred by the program. For instance, the capital cost of a car that already belongs to the organization and that is used in the program’s implementation would be taken into account as part of the full cost, but would not be taken into account as part of incremental cost, 4) Measuring the resources used, 5) Valuing the resources that are used. Capital costing method must be applied to all investment costs (such as those associated with training or social mobilization), 6) Quantifying program outputs, 7) Calculating program costs, taking into account both total and unit costs as well as cost structure in terms of the percentage breakdown for each cost (e.g. capital, labor, material and activity costs). Unit costs are derived by dividing total cost by total quantity of output, and are calculated using full cost and incremental costs, called average cost and marginal cost, respectively.

Whether full costs or incremental costs are used will depend on the proportion of program costs and the total cost of the organization. If program costs make up only a small proportion of the total organization cost, then incremental costs are recommended.

**Analysing the cost of illness and health risks**

**Cost of Illness analysis**

The cost of illness is defined as the economic burden to society caused by a disease or illness. Conducting a cost analysis of a given condition involves five steps: 1) Designing the study-identifying the objective, definition, and scope of the illness, clarifying the approach (whether the focus is on prevalence or incidence), defining the time horizon and perspective, and identifying which type of treatment or health service to examine, 2) Defining the services and resources that correspond to the study design, 3) Measuring the quantity of each component of services and resource that are to be used, 4) Converting the services and resources to a monetary value, 5) Calculating total costs, costs by component, and unit costs, taking into account the various factors that affect the cost.

**Component of costs according to disease progress and treatment**

When calculating illness costs, all of those associated with the major illness and any complications, from first symptom to either death or cure, are included, but costs associated with co-morbidities are not. When adopting the societal perspective, costs also include all kinds of treatment regardless of the health service facilities (if they are associated with the condition) as well as any costs incurred by the patient and family as a result of the condition (transportation, meal, accommodation and other illness-related costs). The societal perspective also takes into account the time...
loss incurred by patients and caregivers (Fig. 1).

Study approach

Two main approaches can be adopted when conducting cost of illness studies-prevalence- and incidence-based approaches(14).

1) Prevalence-based approach

The prevalence-based approach looks at all patients over the time horizon of the study. The time horizon adopted is normally one year to avoid seasonal variation and patients showing symptoms before or during the time horizon. This means that the patients in the study are at different levels of disease progression and severity. In prevalence-based studies, results are presented as cost per person per year (or time horizon).

2) Incidence-based approach

The incidence-based approach looks at all new cases that emerge during a set period of time (normally one year) and follows them until end of the illness (cure or death). Costs are calculated for this period and are known as life time costs. In incidence-based studies, results are presented as cost per episode.

Cost components

Cost components can be categorized in a number of different ways, including medical and non-medical costs. The most common and well-known categorization divides costs into direct medical, direct non-medical and indirect(14).

1) Direct medical costs

Direct medical costs include all of those that are incurred as part of the provision of healthcare—diagnosis, treatment, rehabilitation, and terminal care as well as those costs associated with health services including those incurred as a result of institutional and non-institutional care (e.g. home care or alternative care).

2) Direct non-medical costs

Direct non-medical costs cover all costs borne by patient and family because of receiving healthcare. They include transportation, meal, and accommodation (except for those incurred by in-patients staying in hospital wards, which are already included in direct medical costs), as well as those associated with special devices, home modification, and payment for caregivers. Direct non-medical costs also include those incurred as a result of the time lost by unpaid caregivers (relatives/friends) while the patient is receiving treatment and recovery at home (informal care).

3) Indirect costs

Indirect cost or productivity costs are those incurred as a result of time or working capacity loss caused by illness. Time loss results from a patient’s death or from their inability to attend work or daily activities as a result of the illness, and is subdivided into two periods-whether they are receiving treatment or whether they are in their recovery period. All patient time loss is classified as either a productivity cost or an indirect cost. However, time loss that occurs as a result of a patient receiving treatment may be categorized as a direct non-medical cost(15). In this article, all patient time loss is classified as an indirect cost (Fig. 1). This is to avoid double counting when time loss is measured as an outcome in terms of quality-adjusted life years (QALYs) (for more details, see section Defining cost components).

A well-known method that is used to value time loss is the human-capital method(1). From an ethics standpoint, the value of a specified period of time loss is valued in the same way regardless of the individual’s characteristics or income and regardless of the type of time—whether it is paid work, unpaid work, or leisure(1). The value calculation is made by multiplying the time loss by the rate of income. The rate of income is taken from the per capita Gross National Income (GNI)(16) or the Gross National Product (GNP)(17). Using GNI instead of Gross Domestic Product (GDP) allows the exclusion of productivity data from foreign investment. Per capita GNI is calculated from the total population, and thus includes people of all ages. To calculate morbidity costs,
the cost per day is calculated by dividing the per capita GNI by 365 days. To calculate the cost of informal care, a cost per hour is calculated by adjusting the per capita GNI by 52 weeks per year and 48 working hours per week (according to data from the Thai Department of Labour)(18).

Cost analysis of health risks
The costs associated with health risks are estimated on the basis of their likelihood to lead to particular diseases. For instance, alcohol, drinking, smoking, and obesity can cause a number of leading non-communicable diseases. Firstly, the cost of the illness is measured, then the cost is adjusted by the extent to which the health risk is deemed attributable (known as an attributable fraction), which is calculated from relative risk(19). In addition, costs from the non-health sector might also be included, for instance, damage of assets and crime management from car accidents caused by drink-driving. There are examples of these kinds of study in Thailand that have evaluated the wider cost impact of alcohol drinking(20) and obesity(21).

Defining costs for health economic evaluation
Defining cost components
Cost components that are used in health economic evaluations will vary depending on the perspective adopted and type of the evaluation that is undertaken. There are five major types of perspective: patient, provider, payer, health system, and society. The cost components within a perspective can also vary depending on the type of evaluation and the outcome measure used.

In cost-utility analyses that use the societal perspective, patient quality of life is measured with a QALY (see the relevant article in this volume for more details), which takes into account the effect of time loss in the form of anxiety affecting quality of life and duration of illness. Therefore, no costs associated with patient time loss in terms of indirect costs are included in the analysis, to avoid double counting(15,22). However, some researchers have argued that the QALY does not adequately capture time loss(23) indirect costs may be included in the sensitivity analysis. When conducting a cost-effectiveness analysis using life-years gained as an outcome, the cost components that are included are similar to those that are included in the cost-utility analysis(19). Costs associated with the time loss of caregivers are usually included in the cost-utility analysis, even though they are sometimes defined as indirect costs. This is because time loss of caregivers is not taken into account when measuring patient quality of life using a QALY. Details of the cost components for various types of economic method and perspective are presented in Table 1.

Sources of costing data
The costing data that is used in economic evaluations can come from various sources depending on the objective and application of the study. All studies involved in country policy formulation require nationally representative data. Primary data from appropriate sample size and sampling method are preferred when undertaking economic evaluation, but if that is not available, then cost estimates from standard services using references or standard unit costs can be used. The reliability of secondary data will depend on the quality of studies from which they are derived. Expert opinion is less reliable data and should come from a panel of experts, where possible. Regardless, data from expert opinion should never be a major part of the data used in the analysis.

The most reliable sources of cost data are as follows:
1. Generic drug cost data is used for the base case and data from the original drug is used for the sensitivity analysis. Drug prices are taken from the hospital purchasing prices that are available from the Drug and Medical Supply Information Centre of Ministry of Public Health (http://dmsic.moph.go.th/). Prices are reported by hospitals on a voluntary basis, and the use of a median, garnered from the reported data, is recommended. For analyses that are investigating a proposal submitted to the National List of Essential Drugs or an insurance benefit package, the prices that are quoted in the proposal are used.
2. To measure direct medical costs, data should be only be taken from studies with a valid and appropriate study design(24). First, the number of drugs and medical services used is measured and then the unit cost from the same study site(s) is used in the valuation. If these data are not available, then the standard unit cost of medical services from the standard cost list(12) is used, in conjunction with the aforementioned reference drug prices.
3. All country database input should be derived from resource quantities, adjusted by standard unit cost of medical services, reference drug prices and direct non-medical reference unit cost data (which is available in the standard cost list)(12).
4. Charges from a country database can be
### Table 1. Costs in various perspectives of economic studies

<table>
<thead>
<tr>
<th>Costs</th>
<th>Patient</th>
<th>Provider</th>
<th>Payer</th>
<th>Health system</th>
<th>Society</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of illness</td>
<td>Col</td>
<td>CEA</td>
<td>CUA</td>
<td>CBA</td>
<td>Col</td>
</tr>
<tr>
<td>Direct medical cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital treatment</td>
<td>ch</td>
<td>ch</td>
<td>ch</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Treatment before and</td>
<td>ch</td>
<td>ch</td>
<td>ch</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>after hospital treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct non-medical cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travel, meal and hotel of</td>
<td>ch</td>
<td>ch</td>
<td>ch</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>patient and caregivers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equipment and facilities for</td>
<td>ch</td>
<td>ch</td>
<td>ch</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>patient</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time loss of caregivers</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>x</td>
</tr>
<tr>
<td>Indirect cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time loss of patient due to</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>x</td>
</tr>
<tr>
<td>treatment and recovery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time loss of patient due to</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>x</td>
</tr>
<tr>
<td>death or permanent severe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of healthcare program:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>national immunization program</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direct medical cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost of vaccine</td>
<td>ch</td>
<td>ch</td>
<td>ch</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Vaccine wastage</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>c</td>
</tr>
<tr>
<td>Vaccine logistics</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>e</td>
</tr>
<tr>
<td>(including cold chain)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaccine delivery</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>c</td>
</tr>
<tr>
<td>(including campaign, if any)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management of wastage</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>c</td>
</tr>
<tr>
<td>Adverse events following</td>
<td>ch</td>
<td>ch</td>
<td>ch</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>immunization (AEFI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **c** = cost, **ch** = charge, **i** = income loss, **r** = reimbursement, **x** = not included
- **Col** = cost of illness; **CEA** = cost-effectiveness analysis; **CUA** = cost-utility analysis; **CBA** = cost-benefit analysis
- **CEA** = cost/life saved, cost/life year gained, indirect cost is not included.
- **CUA** = cost/DALY averted, cost/QALY gained
<table>
<thead>
<tr>
<th>Costs</th>
<th>Patient</th>
<th>Provider</th>
<th>Payer</th>
<th>Health system</th>
<th>Society</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Col</td>
<td>CEA</td>
<td>CUA</td>
<td>CBA</td>
<td>Col</td>
</tr>
<tr>
<td>Direct non-medical cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Travel, meal and hotel of patient</td>
<td>ch</td>
<td>ch</td>
<td>ch</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>and caregivers for receiving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vaccine and treatment of AEFI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time loss of caregivers</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>x</td>
</tr>
<tr>
<td>Indirect cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time loss of vaccine receiver</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>x</td>
</tr>
<tr>
<td>due to treatment and AEFI recovery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time loss of vaccine receiver</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>i</td>
<td>x</td>
</tr>
<tr>
<td>due to death or permanent severe</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>disability from AEFI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c = cost, ch = charge, i = income loss, r = reimbursement, x = not included
Col = cost of illness; CEA = cost-effectiveness analysis; CUA = cost-utility analysis; CBA = cost-benefit analysis
CEA; cost/case averted, cost/life saved, for cost/life year gained, indirect cost is not included.
CUA; cost/DALY averted, cost/QALY gained
adjusted to costs using the cost-to-charge ratio available in the standard cost list\textsuperscript{[12]}. These are 1.63 and 1.45 for provincial/regional hospitals and district hospitals, respectively\textsuperscript{[12]}. The cost is a result of charge multiplied by the ratio.

5. Secondary data from past studies must be adjusted to the current study year by applying the consumer price index.

6. Costs may also be estimated from standard practice data, which is derived by applying an appropriate analytical technique (for instance, the Delphi technique) to data from a panel of experts. Drug and medical services used in the guidelines must be converted to costs using reference drug prices and standard cost list referred to in point 1, 2.

**Costing data quality checklist**

The quality of secondary data should also be assessed. Practically, this means assessing the quality of the study design, data collection method, analysis, and results presentation. A quality checklist can be used as a tool to ensure that the quality of the costing study is appropriate. The categories of the checklist are as follows:

1. Perspective
2. Cost composition
3. Year of cost value
4. Sample definition, sample size, data collection method
5. Result presentation composing quantity of resources used, unit cost, total cost and cost classified by composition.

**Costing data adjustment**

To adjust out-of-date costing data, the consumer price index is applied, to ensure that all costs are appropriate for the year of the study\textsuperscript{[20]}. The list below shows which price indices are applied to various costs to ensure they are up to date.

- Drug and medical services: Consumer price index of medical care.
- Transportation: Consumer price index of public transportation services.
- Meals: Consumer price index of food and beverages.
- Accommodation: Consumer price index of shelter.
- Other materials: Consumer price index of all items.

All results should be presented in both local and international currencies. To convert to the international dollar, the purchasing power parity (PPP) exchange rate\textsuperscript{[26]} and the GDP deflator\textsuperscript{[17,27]} should be deployed.

As well as adjustments made to reflect any changes in date and currency, cost data must be also adjusted to reflect the health service facilities that are used in model of analysis, whether that be out-patient service or in-patient service and whether the treatment takes place at a district or provincial hospital. For country-wide analyses, it is recommended that a weighted average of the costs be used.

**Costing data used in health economic evaluation**

As discussed earlier there are several controversial issues surrounding costing for HTA. To help create guidelines that will improve uniformity across studies and improve comparability, the author makes the following recommendations:

1. Costs resulting from patient time loss, incurred both during the treatment and recovery period, should be classified as indirect costs.
2. Time loss costs should be converted to a productivity costs by adjusting them according to Gross National Income, GNI. This corresponds to the data that are used to determine the cost-effectiveness threshold.
3. To calculate the cost of informal care, cost per hour should be used for time conversion. Cost per hour is derived from GNI per capita divided by 52 weeks per year and 48 hours per week (based on 8 hour working days, with 6 working days a week).
4. Informal care costs should be classified as direct non-medical costs and included in the cost-utility analysis. Classification as an indirect cost would result in confusion as indirect costs are not included in the cost-utility analysis.
5. When conducting a cost-utility analysis and cost-effectiveness analysis that measures life-years gained, the time cost of the patient as an indirect cost should not be included. However, it should be included in the sensitivity analysis.
6. Assessment of the reliability of the costing data should be undertaken by assessing the reliability of the source, according to the guidelines detailed herein.
7. The costing data used in health economic evaluations should take into account the following aspects:
   1. Perspective
   2. Source of data (primary or secondary) and details on how it was sourced (including references)
3. Cost components and their values
4. Year of cost value and year of analysis, type of consumer price index used in value adjustment (including exchange rate if appropriate). References should be supplied.
5. The weighted average for nationwide costs should incorporate the cost of each type of health service facility and the proportion of utilization.
6. Data used in the sensitivity analysis and details on how it was sourced.

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

Whether a technology is deemed to be cost effective according to a health economic evaluation depends on both its effectiveness and costs. While effectiveness if often assessed in research, costing is often overlooked. Given this, researchers should pay particular attention to the costing side of this assessment to ensure the costs are adequately captured. In particular, researchers should take care to account for the sources of data, the type of analysis and the manner of data presentation. Although new interventions can be very effective, they may not necessarily be cost-effective. Whether a technology is cost-effective or not can be decided on the basis of solid and detailed costing data together with effectiveness data. An error in calculating costing data can have significant consequences. For instance, an analysis that erroneously finds a certain technology not cost-effective might deprive society of a highly effective—potentially life-saving—technology. Conversely, if a technology is erroneously deemed to be cost-effective and adopted across a society, the costs will outweigh the benefits. Clearly, both cost and effectiveness are essential parts of any HTA.

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

Health Economics; 2002.

การประเมินคุณค่าในการประเมินทางเศรษฐศาสตร์สาธารณะ

อาจารย์วิชิตนุช

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

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 patients has 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
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\(^\text{(}\text{3,4}\text{)}\). In another study, the level of CD4-cell count was not always found to be a good predictor of AIDS-related deaths\(^\text{(}\text{5}\text{)}\).

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 outcomes only to be cited as evidence where there is also a proven association between a surrogate and patient-important outcome\(^\text{(}\text{6}\text{)}\). 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\(^\text{(}\text{7}\text{)}\), 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)\(^\text{(}\text{7}\text{)}\).

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\(^\text{(}\text{6}\text{)}\). 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\(^\text{(}\text{6}\text{)}\).

**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\(^\text{(}\text{6}\text{)}\), 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\(^\text{(}\text{9}\text{)}\) proposed a hierarchy of data sources that are appropriate for use in determining the relative treatment effect. In their study\(^\text{(}\text{9}\text{)}\), 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, 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 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 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 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.
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.

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+ | 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 |

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


การวัดผลผู้ติดสุขภาพ

มาหะบัน อารవิชูยวงศ์

ผลลัพธ์ทางสุขภาพเป็นองค์ประกอบที่สำคัญต้องแน่นอนของการประเมินคุณค่าทางสาธารณสุข ผลลัพธ์ทางสุขภาพที่น่าสนใจในการประเมินคุณค่าทางสุขภาพต้องคำนึงถึงผลประโยชน์ที่ได้จากการวัดผลผู้ติดสุขภาพที่มีการกระจายอยู่ในประชากรที่มีลักษณะใกล้เคียงกับบริบทของการประเมินความรุนแรงแบบวิจัยที่เหมาะสม ซึ่งแบ่งกลุ่มสุขภาพออกเป็นกลุ่มสุขภาพดี การเตรียมการใช้ผลลัพธ์ที่เป็นตัวแทน ในการที่จะเป็นความเห็นอย่างเด่นที่ยังคงและผลักดันความจำเป็นตลอดจนหลักฐานแสดงความเห็นของผลลัพธ์ที่เป็นตัวแทนแล้วกลับไปผลิตผลสุขภาพที่ต้องการ ผลลัพธ์ทางสุขภาพที่น่าสนใจในการประเมินความเป็นข้อมูลปัจจัยสำคัญในการประสานและประสานความแตกต่างของผลผลิตทางสุขภาพที่นำไปใช้ในการวางแผนอื่น ๆ ที่ต้องการเปลี่ยนแปลงการใช้ผลของผลผลิตและทัศนคติที่ทำให้ "ปรับปรุงที่เบี่ยงเบนอาการ" ที่ต้องการศึกษาโดยตรง ทั้งนี้ผลจากการเปลี่ยนแปลงที่ผลกลุ่มสุขภาพสามารถมีผลร่วมกับการพิจารณาใดๆ แล้วว่าจะทำให้ผลผลิตที่ไม่ได้จากการวิเคราะห์เปลี่ยนแปลงทางครอบครัวในกรณีที่มีการเปลี่ยนแปลงที่ควรจะทำให้ผลผลิตที่ดีเพราะจะทำให้ผลผลิตที่ไม่ได้จากการวิเคราะห์เปลี่ยนแปลงทางครอบครัว
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.

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>
<thead>
<tr>
<th>Study</th>
<th>Random sequence allocation</th>
<th>Allocation concealment</th>
<th>Blinding of participants or personnel</th>
<th>Blinding of outcome assessment</th>
<th>Incomplete outcome data</th>
<th>Selective outcome reporting</th>
<th>Other sources of bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study A (2010)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Study B (2011)</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Study C (2011)</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>?</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Study D (2011)</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>?</td>
<td>+</td>
</tr>
</tbody>
</table>

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

+ = 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
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/MA 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 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) and Downs and Black.

The NOS 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. 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 (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.

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. 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:

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

![Fig. 2](image-url) 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,

- $D_{AC-direct}$ represents the relative effects of drug A compared to placebo C.
- $D_{BC-direct}$ represents the relative effects of drug B compared to placebo C.
- $D_{AB-indirect}$ represents the indirect effects of drug A compared to drug B.

Equation 1

$$D_{AB-indirect} = D_{AC-direct} - D_{BC-direct}$$

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 ($D_{AB-direct}$) with indirect evidence ($D_{AB-indirect}$) using a meta-analysis, as shown in Fig. 3B.

Let us suppose the following statements hold:

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

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

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

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

Equation 2

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

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 pairwise 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).
- $\delta^2_{\text{Diff}}$ is variance of the difference of relative effect between direct and indirect evidences (Equation 4).
- $\delta^2_{AB}$ is variance of the relative effect of drug A compared to drug B.
- $\delta^2_{AC}$ is variance of the relative effect of drug A compared to drug C.
- $\delta^2_{BC}$ is variance of the relative effect of drug B compared to drug C.

\[
\text{Diff} = D_{AB\text{-direct}} - D_{AB\text{-indirect}} \quad \text{Equation 3}
\]

\[
\delta^2_{\text{Diff}} = \delta^2_{AB} + \delta^2_{AC} + \delta^2_{BC} \quad \text{Equation 4}
\]

\[
Z_{\text{statistic}} = \frac{\text{Diff}}{\sqrt{\delta^2_{\text{Diff}}}} \quad \text{Equation 5}
\]

**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

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
การพบผู้ป่วยที่มีอาการเป็นระบบและการวิเคราะห์ภูมิคุ้มกันเป็นระบบในการตรวจวัดชั้นสุขภาพ


dr. _#### คุณคุณทุ่น, คุณคุณ เสนแก้ว, รองวิวัฒ ธรรมศิริ, ปิยะทด คีรรรรรรร

การพบผู้ป่วยที่มีอาการเป็นระบบและการวิเคราะห์ภูมิคุ้มกันนับเป็นระบบในระบบการตรวจวัดชั้นสุขภาพ
สำหรับการประเมินโรคในกลุ่มสุขภาพมีการตรวจวัดชั้นสุขภาพของการศึกษา ที่ได้รับการคัดเจาะในการพบผู้ป่วยที่มีอาการเป็นระบบ และการวิเคราะห์
เชิงกิจกรรมและการพบผู้ป่วย อาจเป็นระบบและการวิเคราะห์เชิงกิจกรรมได้รับการพิจารณา หลากหลายแก่เสนอแนวทาง
การประเมินชั้นสุขภาพของการศึกษาให้ได้รับการคัดเจาะในการพบผู้ป่วยที่มีอาการเป็นระบบ และการวิเคราะห์เชิงกิจกรรมและการประเมิน
การพบผู้ป่วยที่มีอาการเป็นระบบและการวิเคราะห์ภูมิคุ้มกัน สำหรับการศึกษาที่รู้จักแบบทดลองโดยมีการรู้และดำเนินการ ผู้พิมพ์แนะนำให้ใช้
"Risk of Bias" ซึ่งเป็นเครื่องมือสำหรับการประเมินคุณภาพการค้นหุ้นการประเมิน ด้านกระบวนการสุข กระบวนการเปิดผลการสุน
การเปรียบเทียบผลของการค้นหุ้นและการรายงานผลการทดลอง สำหรับการศึกษาเชิงสังเกต ผู้พิมพ์แนะนำให้ใช้ "Newcastle-Ottawa scale"
ซึ่งเป็นเครื่องมือที่ประเมินคุณภาพจากข้อมูล 3 ส่วนหลักๆ ได้แก่ การคัดเลือกผู้ป่วยเข้าการศึกษา การเปรียบเทียบของการรักษาที่ได้รับผลลัพธ์
ของการศึกษาหรือการว่าที่ได้รับ อีกฝ่ายมักจะให้ผู้พิมพ์แนะนำให้ใช้สำหรับการประเมินการศึกษาเชิงสังเกต "Down and Black scale"
ซึ่งเป็นเครื่องมือที่ประเมินคุณภาพของการรายงานผลการศึกษา ความเพียงพอของการรายงานผลการศึกษา ด้านการประเมินคุณภาพของการomniaการศึกษาอย่างเป็นระบบและการวิเคราะห์เชิงกิจกรรมเช่น
ผู้พิมพ์แนะนำให้ใช้เครื่องมือ Klassen และคณะ ซึ่งครอบคลุมในส่วนของการตั้งคำถามที่เหมาะสม การคัดเลือกการศึกษา ความครอบคลุมของการศึกษาคุณภาพของการศึกษาที่ได้รับการ
ทบทวนธรรมเนียม ความเหมาะสมในการทำซ้ำ และความเที่ยงตรงของผล 作息นักข่าวเพื่อให้แสดงผลของการวิเคราะห์ภูมิคุ้มกัน ซึ่งควรใช้เครื่องมือ " โดยเฉพาะอย่างยิ่งในกรณีที่มีหลักฐานเชิงประจักษ์โดยตรง หรืออาจมีความคิดการที่จะเปลี่ยนทิศทางผลลัพธ์หลายๆ
ทางเพิ่มอีก ถ้าคือ
Measurement of Utility

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, Bangkok, Thailand

The Quality Adjusted Life Year (QALY) is the most widely recommended health outcome measure for use in economic evaluations. The QALY gives a value to the effect of a given health intervention in terms of both quantity and quality. QALYs are calculated by multiplying the duration of time spent in a given health state, in years, by the quality of life weighted, known as utility. Utility can range from 0 (the worst health state—the equivalent of death) to 1 (the best health state—full health). This paper provides an overview of the various methods that can be used to measure utility and outlines the recommended protocol for measuring utility, as described in the Guidelines for Health Technology Assessment in Thailand (second edition). The recommendations are as follows: Wherever possible, primary data collection using EQ-5D-3L in patients using Thai value sets generated from the general public should be used. Where the EQ-5D-3L is considered inappropriate, other methods such as Standard gamble (SG), Time-trade-off (TTO), Visual analogue scale (VAS), Health Utilities Index (HUI), SF-6D, or Quality of well being (QWB) can be used. However, justification and full details on the chosen instrument should always be provided.

Keywords: Utility, EQ-5D, Quality of life, Quality-adjusted life year

The Quality Adjusted Life Year (QALY) is the most widely recommended health outcome measure for use in economic evaluations(1). The QALY gives a value to the effect of a given health intervention in terms of both quantity and quality. QALYs are calculated by multiplying the duration of time spent in a given health state, in years, by the quality of life weighted, known as utility. Utility can range from 0 (the worst health state—the equivalent of death) to 1 (the best health state—full health). For example, if an individual lives for 10 years with an associate utility of 0.9, this would equal to 9 QALYs (0.9x10). It is also possible to have states regarded as worse than dead, which are represented by a negative value.

Utility value can be derived both directly and indirectly(2). The most common direct methods include Standard gamble (SG), Time-trade-off (TTO) and Visual analogue scale (VAS). Although direct methods offer certain benefits for deriving utility, they can be time consuming and difficult to apply. In contrast, indirect methods that use multi-attribute health status classification systems to derive utility, such as the EuroQol (EQ-5D)(3,4), Health utilities index (HUI)(5), SF-6D(6), and Quality of well being (QWB)(7) are much more convenient and so more widely used. More details on each method can be found in the first edition of HTA guidelines for Thailand(8). It should be noted that the choice of method could significantly affect estimated values for utility(9,10). Thus, it is important that all HTAs define the methodology used as well as the values generated.

According to the International Society for Pharmacoeconomics and Outcomes Research (ISPOR)(1), SG, TTO, and EQ-5D are the three methods that are most widely recommended for measuring utility value in many countries. In the first edition of Thai’s HTA guidelines, EQ-5D-3L was given as the preferred method(11). EQ-5D-3L is a very widely used instrument that is used to measure health outcomes. First developed by the EuroQol group(3,4), the EQ-5D-3L has been translated into 102 official languages, including Thai(2). The EQ-5D questionnaire consists of two parts—the EQ-5D descriptive system and the EQ visual analogue scale (EQ VAS). The EQ-5D descriptive system assesses five dimensions—mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.
Each dimension is assessed by according it one of three levels—no problems (level 1), moderate or some problems (level 2) and severe problems (level 3), giving a total of 243 defined health states that can be measured by the EQ-5D-3L. Each health state can then be described using a five-digit number, where each digit refers to the level of each dimension. For example, 11111, the best health state, refers to a patient who has no noticeable health problems (no problems in walking about, no problems with self-care, no problem with performing usual activities, no pain or discomfort, and not anxious or depressed). On the other hand, 33333, the worst health state, refers to the opposite end of the spectrum (confined to bed, unable to wash or dress without help, unable to perform usual activities, experiences extreme pain or discomfort, and experiences extreme anxiety or depression).

The EQ-VAS is a vertical visual analogue scale, similar to a thermometer, that is used to measure an individual’s assessment of their current health-related quality of life, ranging from 0 (worst imaginable health state) to 100 (best imaginable health state). The Thai EQ-5D-3L can be found in the first edition of HTA guidelines for Thailand(11).

Each health state in the EQ-5D can be converted to a utility value using formula, which essentially attaches weight to each of the levels in each dimension. The formula is based on certain EQ-5D health state values that are calculated from either VAS data and/or TTO data, derived from general population samples. Value sets are then produced for the full set of health states. In Thailand, the value set for EQ-5D-3L health states was derived from a general population representative sample of 1,409 respondents that was taken in 2007, using TTO method(13). The second highest score was 0.766 for state 11112 and the lowest score was -0.454 for state 33333(13).

A utility score for the Thai population can be converted from each EQ-5D-3L health state by subtracting the following from 1: constant terms, dimension-specific coefficients, and N3 terms. The constant term is used to correct for any dysfunction that is present, the dimension-specific coefficient is used to adjust for the level of problems presented in each dimension, and the N3 term is used to make adjustments if any dimension is given a rating of 3. The constant and coefficient values used to calculate the EQ-5D-3L value set for the Thai population are presented in Table 1, and example calculations for health state 11332 and 22222 is shown in Table 2.

In general, the EQ-5D-3L has good psychometric properties(14-17). However, as there are only three levels of response categories, a substantial ceiling effect is usually observed(18,19), which makes it difficult to capture health improvement in respondents whose scores are very high. The EQ-5D-3L, therefore, while useful, is not very sensitive to measuring small changes, especially in mild conditions. To account for this limitation, the EuroQol group has recently launched

### Table 1. Constant and coefficient values used to calculate the EQ-5D-3L value set (adapted from Tongsiri et al)(13)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Level</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td></td>
<td>0.202</td>
</tr>
<tr>
<td>Dimension 1 (mobility)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.432</td>
</tr>
<tr>
<td>Dimension 2 (self-care)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.242</td>
</tr>
<tr>
<td>Dimension 3 (usual activities)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.118</td>
</tr>
<tr>
<td>Dimension 4 (pain/discomfort)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.209</td>
</tr>
<tr>
<td>Dimension 5 (anxiety/depression)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.11</td>
</tr>
<tr>
<td>If answer level 3 in at least 1 dimension (N3)</td>
<td></td>
<td>0.139</td>
</tr>
</tbody>
</table>

### Table 2. Example of calculating utility for health states 11332 and 22222

<table>
<thead>
<tr>
<th>Health state</th>
<th>11332</th>
<th>22222</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting value</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Constant</td>
<td>0.202</td>
<td>0.202</td>
</tr>
<tr>
<td>Dimension 1</td>
<td>0</td>
<td>0.121</td>
</tr>
<tr>
<td>Dimension 2</td>
<td>0</td>
<td>0.121</td>
</tr>
<tr>
<td>Dimension 3</td>
<td>0.118</td>
<td>0.059</td>
</tr>
<tr>
<td>Dimension 4</td>
<td>0.209</td>
<td>0.072</td>
</tr>
<tr>
<td>Dimension 5</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td>N3</td>
<td>0.139</td>
<td>0</td>
</tr>
<tr>
<td>Utility value</td>
<td>0.3</td>
<td>0.393</td>
</tr>
</tbody>
</table>
the EQ-5D-5L, which has the same five dimensions as the original EQ-5D-3L, but with five levels on each dimension rather than three\textsuperscript{20}. To date, 91 language versions of the EQ-5D-5L have been produced, including Thai\textsuperscript{22}. The EQ-5D-5L has been validating in several groups of patients across six countries. It was found to have greater discriminative power than the EQ-5D-3L and to exhibit less ceiling effect. The EQ-5D-5L was also found to have higher convergent validity when compared to the EQ-5D-3L\textsuperscript{21,22}. To date, no Thai value set derived from Thai population data has yet been produced for the EQ-5D-5L. However, Thailand’s Health Intervention and Technology Assessment Program (HITAP) is currently undertaking research to establish value sets for the EQ-5D-5L using the EQ-VT protocol developed by the EuroQol group. The study is expected to be completed in 2014. In the meantime, the Euroqol group has developed crosswalk value sets for EQ-5D-5L in a number of countries, including Thailand, Denmark, France, Germany, Japan, the Netherlands, Spain, UK, US, and Zimbabwe. These crosswalk sets allow the value sets from the EQ-5D-3L to be used in EQ-5D-5L, and were developed using a response mapping approach\textsuperscript{23}. As no value set derived from the Thai population exists, EQ-5D-5L is currently the preferred method used to measure utility.

Evaluation of health state differs according to whether patients or the general population are being surveyed. There is some debate as to whether it is more appropriate to survey the general population or patients living with the condition, when measuring utility for HTAs. In general, patients tend to give health states higher values than do the general population\textsuperscript{24-27}. This is probably due to a number of factors, including adaptation to the health state and changes in perception resulting from living with a certain condition\textsuperscript{28}. The general population may also focus disproportionately on the negative aspects of a given health state\textsuperscript{28,29}. Moreover, patient value may be a more accurate measure, as patients know the reality of a given health state better than the general population, who have to imagine what it is like to live with the condition, rather than base their responses on genuine experience. Nevertheless, the Washington Panel on Cost-effectiveness in Health and Medicine as well as HTA guidelines from many countries\textsuperscript{31,32} still recommended basing utility on values derived from the general population rather than a patient data set. The reason for this preference is two-fold—first, that tax-funded health care systems should be based on values that reflect the preferences of all taxpayers\textsuperscript{33,34}, and second that general population surveys generate more neutral values, as self-interest is less likely to influence their responses (the so-called “veil of ignorance” theory)\textsuperscript{32}. As such, garnering utility data from general population samples, rather than patients, seems to be more appropriate for HTAs.

In the UK, the National Institute of Health and Clinical Excellence (NICE)\textsuperscript{35} recommends that “For the reference case, the measurement of changes in HRQL should be reported directly from patients, and the value of changes in patients’ HRQL (that is, utilities) should be based on public preferences, using a choice-based method.” The EQ-5D is the preferred measure of HRQL in adults. The methods to elicit EQ-5D utility values should be fully described. When EQ-5D data are not available or are inappropriate for the condition or effects of treatment, the valuation methods should be fully described and comparable to those used for the EQ-5D. Data collected using condition-specific, preference-based measures may be presented in separate analyses. The use of utility estimates from published literature must be supported by evidence that demonstrates that they have been identified and selected systematically\textsuperscript{35}.

The Australian guidelines of the Pharmaceutical Benefits Advisory Committee (PBAC)\textsuperscript{36} recommend indirect method including the Health Utilities Index (HUI2 or HUI3), the EQ-5D, SF-6D, or the Assessment of Quality of Life (AQoL). However, none of these is singled out as the optimum measurement, as all have their own strength and weakness. The use of any other instruments should be justified.

Guidelines for pharmacoeconomic research in the Netherlands\textsuperscript{37} recommend that, in the case of empirical studies, EQ-5D or HUI2/3 that are completed by patients or by proxy can be used. In the case of modelling, two approaches are recommended. The most appropriate approach is to value health states by conducting a representative sample survey from the population, using one of the following methods: standard gamble, time trade-off, or visual analogue scale. If this approach is not possible, or not deemed appropriate, the next recommended approach is to take utilities from published studies. Where different utilities are taken from different sources, the method of assessment, type of assessor, and context of the assessment task should be similar.

Canadian guidelines for HTA assessment\textsuperscript{38} stipulate that both indirect and direct methods can be
used to measure utility. However, the method should be justified fully, and a detailed explanation of its validity and reliability should be provided. When treatment has significant impact on the patient’s caregiver, the quality of life of the caregiver should also be measured. However, this should be reported separately in the analysis and excluded when calculating the Incremental Cost Effectiveness Ratio (ICER).

According to NICE guidelines\(^{(35)}\), the EQ-5D is the recommended instrument for measuring utility in adults. However, measuring utility among children is more complex and challenging due to the lack of appropriate instruments for use in children, especially those who are younger than 5 years old\(^{(39,40)}\). The direct method, for instance, is unsuitable for use in children due to their undeveloped cognitive abilities\(^{(41,42)}\). At present, a few instruments that use the indirect method have been developed for use in children. The EQ-5D, the recommended instrument in adults, is suitable for children aged 12 years old or older and the EQ-5D-Y was developed for use in children aged between 7 to 12 years old. However, no current value sets for the EQ-5D-Y have yet been developed. The HUI2, which NICE recommends as an instrument for measuring utility in children, also requires the use of some form of proxy for very young children (5 to 8 years old)\(^{(44)}\). In addition, a specific value set for HUI2 has been developed for the UK population but not for the Thai population yet.

A number of studies have revealed the significant variety in the range of utility values used in HTA. For instance, one review of the utility values used in a cost effectiveness model of treatment for osteoporosis patients ranged from 0.28 to 0.72 for hip fractures and from 0.31 to 0.8 for vertebral fractures\(^{(45)}\). This wide range may be due to the large number of instruments that are available, as well as differences in the age, social background, and nationality of those surveyed and the severity of their condition\(^{(46)}\). As a result, when utility values are derived from published literature, clear justification for the use of the data should be provided, especially in terms of similarity of source population and appropriateness of the instruments. In their 2005 study, Cooper et al.\(^{(47)}\) proposed a hierarchy of data sources that can be used to derive utility. In their studies\(^{(47)}\), either 1) direct utility assessment for the specific study from a sample either a) of the general population, b) with knowledge of the disease(s) of interest, and c) of patients with the disease(s) of interest or 2) indirect utility assessment from specific study from patient sample with disease(s) of interest, using a tool validated for the patient population were at the top of the hierarchy. On the other hand, expert opinion was at the lowest of the hierarchy.

**Guidelines for Health Technology Assessment in Thailand (second edition): Recommendations for measuring utility value**

1. A set of utility values should be derived from general population data.

2. The EQ-5D-3L is the preferred instrument for deriving utility value, due to the validity, reliability, responsiveness, feasibility and availability of the established value set for Thai population. Primary data collection in patients with similar characteristics should be conducted, using the EQ-5D-3L value set derived from the Thai general population. Appropriate sample sizes should also be determined.

2.1 In some conditions, where the use of EQ-5D-3L is considered inappropriate, alternative methods may be used to measure utility. Direct methods that may be suitable include SG, TTO, and VAS in the general population. Indirect methods that may be appropriate include standardised and acceptable preference-based questionnaires, such as the HUI, SF-6D, or QWB. However, in all cases where the EQ-5D is not used, full justification of the method choice should be provided, and the selected method should be described in full.

2.2 When EQ-5D data is not available, the utility data that is derived from other Health Related Quality of Life (HRQOL) instruments can be used and mapped onto the EQ-5D instrument. In all cases, full justification should be provided and the mapping function should be validated and fully described.

2.3 For children, the recommendations are as follows:

   - For children aged 12 years or older, the EQ-5D-3L is the preferred instrument for measuring utility. Where the EQ-5D-3L is used with children aged 12 years or older, all changes in HRQOL data should be derived from the direct reports of the children, and the value of the change in the children’s quality of life (utility) should be based on general population preference, derived from the Thai EQ-5D-3L value set.

   - For children aged between 8 and 12 years old, the EQ-5D-3L can also be used. Where the EQ-5D-3L is used with children aged between 8 and 12 years old, change in HRQOL should be reported by proxy, and the value of the change in the children’s quality of life (utility) should be based on general population preference.
preference, derived from the Thai EQ-5D-3L value set. Alternatively, any HUI2 data that is derived from direct reports of children can be used in conjunction with the UK value set. Justification should be provided.

- For children aged 8 years old or younger, the EQ-5D-3L can also be used. Where the EQ-5D-3L is used with children aged 8 years or younger, change in HRQOL should be reported by proxy, and the value of the change in the children’s quality of life (utility) should be based on general population preference, derived from the Thai EQ-5D-3L value set. Alternatively, any HUI2 data that are derived from direct reports of children can be used in conjunction the UK value set. Justification should be provided.

2.4 Where utility values are obtained from the published literature, the methods that have been used to garner the data should be systematic and transparent. Utility values should be derived from the Thai population or a population that has similar characteristics to the Thai population. The method used to measure utility should comply with the strictures outlined in the Thai’s HTA guidelines and should be described in full.

2.5 Reliance on expert opinion should be avoided.

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

16. Janssen MF, Lubetkin EI, Sekhobo JP, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes


การวัดความประณีต

มัณฑารัตน์ อาวุธวิทยาพงษ์

ปัจจุบัน (QALY) เป็นผลผลิตทางสุขภาพที่ได้รับค่าความนาในในการประเมินทางเศรษฐศาสตร์ทางภาวะการที่ดี โดยปัจจุบันเป็นการวัดผลของมาตรการทางสุขภาพทั้งในเชิงปริมาณและในเชิงคุณภาพ ปัจจุบันสามารถคำนวณได้โดยการวัดผลทางสุขภาพที่ดีในสถานะสุขภาพนั้นๆ คิดเป็นปัจจุบันความประณีต ซึ่งเป็นค่าความนาที่ดีที่สุดและคุณภาพชีวิต ที่ดีที่สุดประณีตจะมีค่าดังเด่น 0 (ภาวะรุษาพบที่ย่างที่สุดหรือที่จะเสียสุขภาพสูงสุด) บทความนี้จะอธิบายถึงการดังๆ ที่ใช้ในการวัดค่าต่อการปรับประณีต ตลอดจนสรุปข้อแนะนำสำหรับการวัดค่าประณีตชนคล้ายๆ ในคู่มือการประเมินผลทางสุขภาพส่วนบุคคลของประเทศไทย ฉบับที่ 2 ที่มีการแนะนำโดยที่ดีที่สุดเมื่อเป็นไปตามแนวทางการวัดค่าประณีตตามสมมุติฐานประณีต EQ-5D-3L และได้ตั้งประเด็นอธิบาย.WARNING: This text may not be entirely accurate or complete.

Standard gamble (SG), Time-trade-off (TTO), Visual analogue scale (VAS), Health Utilities Index (HUI), SF-6D หรือ Quality of Well Being (QWB) ได้ อย่างไรก็ตามควรแสดงผลด้านการประเมินผลเฉพาะของเครื่องมือดังกล่าวอย่างครบถ้วน

J Med Assoc Thai Vol. 97 Suppl. 5 2014
Handling Time in Economic Evaluation Studies

Unchalee Permsuwan PhD*,
Kansinee Guntawongwan MA**, Piyaluk Buddhawongsa PhD**

* Department of Pharmaceutical Care, Faculty of Pharmacy, Chiang Mai University, Chiang Mai, Thailand
** Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand

The discount rates and time horizons used in a health technology assessment (HTA) can have a significant impact on the results, and thus the prioritization of technologies. Therefore, it is important that clear guidance be provided on the appropriate discount rates for cost and health effect and appropriate time horizons. In this paper, we conduct a review of relevant case studies and guidelines and provide guidance for all researchers conducting economic evaluations of health technologies in the Thai context. A uniform discount rate of 3% is recommended for both costs and health effects in base case analyses. A sensitivity analysis should also be conducted, with a discount range of 0-6%. For technologies where the effects are likely to sustain for at least 30 years, a rate of 4% for costs and 2% for health effects is recommended. The time horizon should be long enough to capture the full costs and effects of the programs.

Keywords: Time horizon, Discounting, Economic evaluation guideline, Thailand

J Med Assoc Thai 2014; 97 (Suppl. 5): S50-S58
Full text, e-Journal: http://www.jmatonline.com

Time horizon refers to the time period over which cost and outcome data can be measured. The length of a time horizon will vary depending on several factors, including the nature of the disease, the budget, the population, and the nature of the research. The time horizon should be designed so that is long enough to capture all relevant costs and outcomes, and to ensure that the results are useful, accurate, and relevant. In practice, short time horizons are often favored in study designs, as they are regarded as more practical and less expensive. However studies that use longer time horizons often result in economic evaluation data that are much closer to reality, as they allow the researcher to monitor the long-term consequences of the health technology under investigation, including those associated with medicinal side effects or drug resistance, which often manifest themselves later.

As both long and short time horizons offer various but differing benefits, choosing which time horizon should be used in any Health Technology Assessment (HTA) is often complex and challenging. One unresolved issue related to the choice of time horizon is how best to incorporate the effect of interventions in complex scenarios, such as those associated with various non-communicable diseases (for instance, how best to capture the effect of cardiovascular therapy on diabetes). It is known, for instance, that many interventions that extend life may result in future unrelated costs and may have no health-related effect on the patient beyond those associated with the natural aging process. The US Public Health Service Panel on Cost-Effectiveness in Health and Medicine recommends that individual researchers use their own judgment when deciding whether to include or exclude these costs and benefits. If the costs are small relative to the magnitude of the cost-effectiveness ratio, they can be excluded. On the other hand, if the costs are large relative to the magnitude of the cost-effectiveness ratio, they recommend using a sensitivity analysis to assess the effect of these costs and benefits.

The WHO recommends that a Cost-Effectiveness Analysis (CEA) be used to evaluate all interventions over a period of 10 years at full implementation. However, this time horizon might not be appropriate in some situations, especially when evaluating vaccinations or treatments for chronic diseases, where the time horizon for the analysis clearly needs to be longer. Analyses must capture all of the health effects of the intervention that occur during the 10-year time horizon as well as those that occur subsequently. A general rule, which has been validated by a recent study review that compares different guidelines from various countries, is that the time...
The theory behind discounting

For one-year projects, the net benefit, cost-benefit ratio, or cost-effectiveness ratio can be easily calculated, and the results of various alternatives can be compared. However, many projects continue for more than one year, and sometimes the costs and benefits of a project can occur more than a year after treatment. This makes it hard to compare the costs and benefits of treatments because their values may vary according to different time periods. In order to make costs and benefits compatible, all values should be adjusted to present values (present worth), and future values should be adjusted by a certain rate called “discount rate”.

An overview of how discount rates work, and a summary of two well-known economic concepts related to discount rate—time preference and opportunity cost of capital—are explained in detail in the first edition of the HTA guidelines for Thailand.

Should discount rate be equal for both costs and effects?

Discounting is performed to adjust future costs and effects for their differential timing. This helps decision-makers compare costs and effects for the same point of time. It is a common practice in health economic evaluations to perform discounting on both future costs and benefits.Discounting future costs and benefit in cost-benefit analyses (CBAs) is widely accepted. However, the practice of discounting for life years saved and quality-adjusted life years (QALYs) in cost-effectiveness analyses (CEAs) and cost-utility analyses (CUAs) has become controversial in recent years. One reason for this is that health, unlike wealth, cannot be invested to produce future gains, which led some scholars to suggest that health effects should be discounted at a very low rate (1.5-2%) or even not discounted at all. A detailed analysis of the advantages and disadvantages of utilising a uniform or differential discount rate in economic evaluations is given in the first edition of HTA guidelines for Thailand.

National Institute for Health and Clinical Excellence (NICE) guidance

Many countries follow the healthcare project, appraisal guidance of the UK’s National Institute for Health and Clinical Excellence (NICE). In its first recommendation of this kind, published in 2001, NICE recommended that health benefits be discounted at a lower rate than costs (1.5% for health benefits as opposed to 6% for costs), in line with the recommendation of the UK Department of Health. The rationale behind using different discount rates was to adjust for the increasing value of health effects. When health effects were measured in quantities such as QALYs, discounting the health effects at a lower rate than that used to adjust for costs enabled the analysis to take into account any increase in the future value of health effects. In 2004, NICE revised the guideline by requiring that both costs and effects be discounted at a 3.5% rate, the social time preference rate (STPR) stipulated by the UK Treasury. This recommendation remains in use today, despite a suggestion, made in 2011, that NICE use a lower discount rate for health effects than costs. This was considered briefly as a way to capture accurate data from projects where the treatment cost is borne immediately and the health effects are felt far into the future, data that studies following the current guideline often missed. For example, the treatment of bone cancer in children and young people was originally deemed not cost-effective, based on the NICE threshold. However, the suggested revision of the guideline was not accepted; instead, a smaller modification was made in the form of suggesting that an additional sensitivity analysis be conducted when the treatment effects are both substantial in restoring health and sustained over a very long period (at least 30 years).

Shortly after NICE stipulated that both health effects and costs be discounted at the same rate, a number of researchers challenged this recommendation, instead claiming that the use of different discount rates was more appropriate. In economics, the policy objective of any government agency is the optimisation of social welfare within the constraints of the budget. This objective determines the criteria for the optimal rates that should be used when discounting costs and health effects. It is clear, therefore, that government agencies should apply a lower discount rate for health effects than they do for costs, as the consumption value of health effects and the cost-effectiveness threshold increase overtime.

Claxton et al have suggested that the social objectives of the health-care decision-makers for whom the CEA is being conducted should determine whether change in the cost-effectiveness threshold and the consumption value of health overtime be considered or not. These objectives are defined by either, 1) present net consumption value of health.
maximisation, which is derived by monetising both the costs and the effects of the health technology or 2) present net value of health maximisation, which is derived by measuring the social welfare generated by the health technology. With the former objective, health technology is evaluated by monetising both costs and effects. The net health effect, given in terms of monetary value represents a measure of social welfare in economics, therefore guaranteeing that the technology that maximizes net monetary value will be chosen. In contrast, with the latter objective, health effects are measured in quantity rather than in monetary value, by comparing the incremental cost-effectiveness ratio (ICER) value with a cost-effectiveness threshold. An optimal health technology is one for which the ICER is less than the threshold. Nevertheless, satisfaction of this objective does not necessarily mean that the technology maximizes social welfare, because ICER is only used to prioritise health technologies. Therefore, the technology that maximizes net monetary value should be chosen.

For this reason, when assessing health effects purely in terms of monetary benefit, any change in the consumption value of health effects and cost-effectiveness threshold are fully accounted for, making further adjustment unnecessary. In contrast, when assessing health effects in terms of the social benefit, health effects are not monetized and, therefore, some adjustments must be made to handle the changes in both values. One practical way to incorporate such changes is to modify the discount rates for costs and health effects.

Claxton et al support NICE’s recommendation that an STPR of 3.5% be used for consumption or social time preference rates, as the discount rates for costs and health effects are optimal only under some conditions. The appropriate discount rates vary according to the decision rule, which is related to the social objective. Table 1 shows the optimal discount rates for health effects and costs for both social objectives. Where the social objective is the maximising of welfare (net present consumption value of health), the health effects discount rate \( d_h \) is approximately equal to \( r_c - g_v \) less the growth rate of health value \( g_v \). This is because when \( g_v \) is positive, the value of health consumption will be valued by society more highly in the future, and \( d_h \) should be less than \( r_c \). However, the discount rate for the costs \( d_c \) is \( r_c \) minus \( g_v \), because the health for gone by the adoption of the technology will also be valued more highly in the future. In addition, the discount rate for the costs must include the growth in the cost-effectiveness threshold \( g_k \). This adjustment is necessary because future costs will displace less future health if the threshold increases over time. If technology is fixed but the health budget increases overtime, the positive value of \( g_k \) implies that the technology adopted in the future will be less cost-effective compared to that adopted in the present. It is clear, therefore, that an increase in health value does not justify the use of different discount rates. However, it does indicate that the use of a rate below the social time preference rate for both discount rates is fitting. The gap between the discount rates for costs and health effects depends on the growth rate of the threshold only. For the maximization of this welfare objective, therefore, NICE guidance should only be adopted where there is no growth in either the cost-effectiveness threshold or the health value \( g_k = 0 \) and \( g_v = 0 \).

Where the social objective is the maximizing of the present net value of health, health benefits are measured using instruments such as QALYs. In this case, the discount rate for health effects \( d_h \) is approximately equal to the social time preference rate for health \( r_h \), while the discount rate for costs \( d_c \) is \( r_h + g_k \). If the cost-effectiveness threshold increases over time, then future cost becomes less important. Under this social objective, NICE guidance is appropriated only where there is no growth in the cost-

<table>
<thead>
<tr>
<th>Social objectives</th>
<th>Discount rate for health ( d_h )</th>
<th>Discount rate for cost ( d_c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Net present consumption value of health maximisation</td>
<td>( -r_c - g_v )</td>
<td>( -r_c - g_v + g_k )</td>
</tr>
<tr>
<td>2. Net present value of health maximisation</td>
<td>( -r_h )</td>
<td>( -r_h + g_k )</td>
</tr>
</tbody>
</table>

Claxton K., et al. (2022)

d\(_h\) = Discount rate applied to unadjusted health gains; d\(_c\) = Discount rate applied to unadjusted costs; g\(_k\) = Growth rate of the threshold; r\(_c\) = Social time preference rate for health; g\(_v\) = Growth rate of consumption value of health; r\(_c\) = Social time preference rate for consumption

---

Table 1. Optimal discount rates for health effects and costs under different social objectives

---

S52 J Med Assoc Thai Vol. 97 Suppl. 5 2014
effectiveness threshold ($g_k = 0$), and the social time preference rate for health is equivalent to the social time preference rate for consumption ($r_h = r_c$).

When using the same discounting rate for costs and health effects (as recommended by NICE), social optimality is only obtained when the cost-effectiveness threshold is constant over time. Theoretically, two factors affect a cost-effectiveness threshold: 1) expectation about health-care budget and 2) health technology development. The threshold will increase if the budget is expected to increase and it will decrease if the technology becomes more cost-effective. Therefore, fixing the cost-effectiveness threshold constant over time is an acceptable assumption. However, when policy makers adjust the threshold, it will decrease if the technology becomes more cost-effective. As Table 2 shows, given a 30-year time horizon, conducting a pap smear every 5 years (in patients aged 30-60 years), generated an ICER of 199,942, 239,032, and 365,388 THB/QALY when a discount rates of 3%, 5%, and 10% were used, respectively.

4. Applying a different discount rate to costs and health effects study findings on the efficiency of the health technology:

- When the health effect is discounted at a higher rate than cost, ICER goes up as the gap between the discount rate of the health effect and cost increases. For example, in studies with a 30-year time horizon, where the health effect of a pap smear every 5 years (for patients aged 30-60 years) was discounted at 3%, 5%, and 10%, and cost was not discounted, the ICER was equal to 326,885, 539,680, and 1,805,551 THB/QALY, respectively.

- When health effect is discounted at a lower rate than cost, the ICER goes down as the gap between the discount rate of the health effect and cost increases. For example, within a 30-year time horizon, when health effect was not discounted but the cost of conducting a pap smear every 5 years (for patients aged 30-60 years) was discounted at 3%, 5%, and 10%, and cost was not discounted, the ICER equal to 326,885, 539,680, and 1,805,551 THB/QALY, respectively.

- When health effect is discounted at a lower rate than cost, the time horizon within which the health technology becomes cost effective becomes shorter. For example, with a 30-year time horizon, where the health effect of conducting a pap smear every 5 years (in patients aged 30-60 years) was not discounted, but the cost was (at a rate of 3%), the ICER was found to fall under the threshold of 92,232 compared with 160,000 THB/QALY. If health effects and costs were both discounted at 3%, the ICER was found to be equal to 199,942 THB/QALY, rising to 326,885 THB/QALY when health effect was discounted at 3% and cost was not discounted.

This finding corroborates the findings of Bos et al(28), who reviewed the impact of discounting for vaccination programs, diabetes interventions, and cancer interventions, and showed that in preventive...
Table 2. The impact of discount rate and time horizon upon Incremental Cost-Effectiveness Ratio

<table>
<thead>
<tr>
<th>Technology</th>
<th>Incremental Cost-Effectiveness Ratio (ICER)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cost 0% 3% 3% 0% 5% 5% 0% 10% 10% 0%</td>
</tr>
<tr>
<td></td>
<td>Effect 0% 3% 0% 3% 5% 0% 5% 10% 0% 10%</td>
</tr>
<tr>
<td>Pap smear every 5 years (age 30-60)</td>
<td></td>
</tr>
<tr>
<td>20 years</td>
<td>6,221,664 9,218,510 3,972,326 14,438,512 12,952,865 2,966,938 27,162,136 105,015,895 1,465,128 445,949,745</td>
</tr>
<tr>
<td>30 years</td>
<td>150,790 199,942 92,232 326,885 239,032 66,787 539,680 365,388 30,515 1,805,551</td>
</tr>
<tr>
<td>Lifetime</td>
<td>-68,215 -59,714 -12,704 -320,642 -4,009 -815,608 12,139 133 -6,235,762</td>
</tr>
<tr>
<td>VIA every 5 years (age 30-45)</td>
<td></td>
</tr>
<tr>
<td>20 year</td>
<td>4,369,407 6,260,609 2,789,687 9,805,815 8,447,854 2,083,506 17,716,350 34,575,728 1,028,596 146,875,396</td>
</tr>
<tr>
<td>30 year</td>
<td>67,701 101,993 47,092 146,628 129,181 36,155 241,897 216,586 18,165 807,205</td>
</tr>
<tr>
<td>Lifetime</td>
<td>-74,354 -72,471 -16,836 -320,058 -6,385 -768,327 -25,629 -367 -5,193,478</td>
</tr>
<tr>
<td>VIA every 5 years (age 30-45) + Pap smear every 5 years (age 50-60)</td>
<td></td>
</tr>
<tr>
<td>20 years</td>
<td>4,369,407 6,260,609 2,789,687 9,805,815 8,447,854 2,083,506 17,716,350 34,575,728 1,028,596 146,875,396</td>
</tr>
<tr>
<td>30 years</td>
<td>67,701 101,993 47,092 146,628 129,181 36,155 241,897 216,586 18,165 807,205</td>
</tr>
<tr>
<td>Lifetime</td>
<td>-72,842 -69,031 -14,863 -338,316 -59,271 -885,872 -18,288 -208 -6,397,858</td>
</tr>
</tbody>
</table>

Results provided only some examples.; *ICER negative meaning cost saving (cost reduced with better health effect); Base case scenario included 20% coverage of screening interventions, 100% coverage of human papillomavirus (HPV) vaccine, lifetime duration of vaccine protection and 15,000 THB for full immunization.
programs with distant future health gains, such as infant vaccination programs or certain screening programs, discounting health effects has a strong impact on the results of the cost-effectiveness analyses. In many studies, the impact is so large that the discount rate is likely to influence decision-making.

Comparisons of the international economic evaluation guidelines

The International Society of Pharmacoeconomics and Outcomes Research (ISPOR) compiles country-specific guidelines for economic assessment, using the criteria for economic assessment as defined by Hjelmgren et al., from around the world (they currently list 35 sets of guidelines from 34 countries around the world). ISPOR divides the guidelines into the following three types:

1) Published Pharmacoeconomic Recommendations: Country-specific economic evaluation guidelines or recommendations published by experts in the field but, not officially recognised or required by the health care decision-making bodies/entities in the country/region for reimbursement.

2) Pharmacoeconomic Guidelines: Country-specific official guidelines or policies concerning economic evaluation that are recognised or required by the health care decision-making bodies/entities in the country/region for reimbursement.

3) Submission Guidelines: country-specific, official guidelines or policies concerning drug submission requirements that include stipulations on economic evaluation must be adhered to by the health care decision-making bodies/entities in the country/region for reimbursement. Whereas most of the guidelines recommend that the time horizon of an economic evaluation should be long enough to capture all the essential costs and health effects, most do not stipulate a specific length of time. Four countries—Italy, Russia, Switzerland and Israel—do not include any information at all on appropriate time horizons for economic evaluations.

Most of the guidelines collected by ISPOR, stipulate the use of the same discount rate for costs and health effects, and many give specific rates. The average identical discount rate in the ISPOR selection was 4.32%, and the minimum rate and maximum rate were 3% and 7%, respectively (Table 3). Some countries do not specify a specific discount rate, and they were not included in this part of the analysis. For example, China recommends using a one-year interest rate for both costs and effects, whereas France, Switzerland, and Finland suggest the use of multiple equal discount rates. Guidelines from Belgium, the Netherlands, Poland, and Scotland all suggest that health effects should be set at a lower rate than that used for costs (which is consistent with the recommendation of British Medical Journal).

Our analysis show that equal discount rates for costs and effects should be used where the cost-effectiveness threshold is constant; a higher discount rate should be set for costs when the threshold is expected to increases overtime. The World Health Organisation (WHO) suggests that the threshold that should be used in CEAs of health care technologies should never be greater than three times the value of per capita Gross Domestic Product (GDP), which reflects the assumption that the GDP per capita tends to rise over time and thus, too will the threshold. In the

Table 3. Discount rates for costs and health effects from international guidelines

<table>
<thead>
<tr>
<th>Discount rate</th>
<th>Number of guidelines</th>
<th>Discount rate (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Same discount rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single rate</td>
<td>19</td>
<td>4.32</td>
</tr>
<tr>
<td>Multiple rate</td>
<td>3</td>
<td>3.56</td>
</tr>
<tr>
<td>Others</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Differential discount rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costs</td>
<td>5</td>
<td>4.4</td>
</tr>
<tr>
<td>Effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Not specify</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

ISPOR(29)
case of Thailand, in 2007, the cost-effectiveness threshold was initially set at 100,000 baht per QALY\(^{(30)}\). Later, the threshold was revised to 160,000 baht per QALY, where it has remained ever since\(^{(31)}\), indicating that the threshold has not increased over time in Thailand.

**Guidelines for Health Technology Assessment in Thailand (Second Edition): Recommendations for handling time in economic evaluations**

**Time horizon**

The time horizon used in any economic evaluation of health technologies in Thailand should be long enough to capture the full costs and effects of the intervention. Researchers are encouraged to use modelling techniques and/or epidemiologic data to estimate future costs and effectiveness and then deploy appropriate discounting.

**Discounting**

Where appropriate, costs and health effects should be discounted at the same annual discount rate—3%, a sensitivity analysis should also be conducted using a uniform discount rate for costs and health effects ranging from 0-6%. The finding should be displayed as a tornado diagram. Where effects are substantial and exhibit over a very long period (30 years or more), a rate of 4% for cost and 2% for health effect should be applied, so that the cost-effectiveness threshold changes in line with the growth of health value \((g_v = g_e)\). This ensures a difference between the two rates of 2%, which is in line with recommendations of other international guidelines and NICE.

**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 National Health Security Office, the Thailand Research Fund under the Senior Research Scholar on Health Technology Assessment (RTA5580 010) and Thai Health Global Link Initiative Program (TGLIP), supported by Thai Health Promotion Foundation.

**Potential conflicts of interest**

None.

**References**

การจัดการกับเอดในภาวะเหมาค่าทางสาธารณสุข

อ.ไชยชัย เพ็ญสุวรรณ, ท่านติชัย ถิ่นทองคำ, ปัชราภรณ์ พุทธวงศ์

กรมเวาและการใช้อัตราอัตราร้อยหรือแต่ละกลุ่มวัยปรับลดลงทุนและผลลัพธ์เป็นประเด็นสำคัญในการประเมินคุณภาพกลุ่มค่าทางสาธารณสุขซึ่งมีผลกระทบของการจัดตั้งค่าสูงสุดของการตีบล่าขุนที่ตีบเหลือกัน นักวิชาการได้นำเสนอหุ้นและผลลัพธ์ของการให้ใช้ความคุ้นหรือจัดการผ่านในประเทศไทย ในการสนับสนุนการลดลงทุน และผลลัพธ์ที่ใช้ รวมทั้งความเป็นมาตรฐานในการประเมินคุณภาพการแพทย์ในประเทศไทย ออกผลที่มีการนั่นๆ ที่แสดงทางเดินพื้นที่ 30 ปีขึ้นไปที่การวิเคราะห์ความไวที่อัตราการลดลง 4 สำหรับขุนที่และยังจะ 2 สำหรับผลลัพธ์
Sensitivity Analysis for Handling Uncertainty in an Economic Evaluation

Supon Limwattananon MPHM, PhD*

*Faculty of Pharmaceutical Sciences, Khon Kaen University, Khon Kaen, Thailand

To meet updated international standards, this paper revises the previous Thai guidelines for conducting sensitivity analyses as part of the decision analysis model for health technology assessment. It recommends both deterministic and probabilistic sensitivity analyses to handle uncertainty of the model parameters, which are best represented graphically. Two new methodological issues are introduced—a threshold analysis of medicines’ unit prices for fulfilling the National Lists of Essential Medicines’ requirements and the expected value of information for delaying decision-making in contexts where there are high levels of uncertainty. Further research is recommended where parameter uncertainty is significant and where the cost of conducting the research is not prohibitive.

Keywords: Acceptability curve, Expected value of information, Sensitivity analysis, Threshold analysis, Tornado diagram

This paper gives an overview of the key methods used to conduct sensitivity analyses (SA) in economic evaluations of health technologies and therapeutic interventions. It revises previous recommendations, published in 2008(1), by addressing and updating several methodological issues that have been highlighted since then by the National Institute for Health and Clinical Excellence (NICE) Guide to the Methods of Technology Appraisal(2), the International Society for Pharmaco-economics and Outcomes Research (ISPOR), and the Society for Medical Decision Making (SMDM) Good Research Practices in Modelling(3).

This paper focuses on approaches to handling parameter uncertainty, a second-order uncertainty in decision analysis models. Issues arising from first-order uncertainty due to random error and structural or model uncertainty due to assumptions are not covered. This article provides a number of recommendations, with concrete examples on the appropriate graphical presentation of results from deterministic(4) and probabilistic sensitivity analyses(5,6). Additionally, a threshold analysis of medicines’ unit prices is recommended as a way to fulfill the cost-effectiveness criteria set by the committee on National Lists of Essential Medicines (NLEM). Lastly, the concept of expected value of information is introduced as a way to determine whether delaying decision-making at the expense of further research to avoid the parameter uncertainty is a valid decision.

Deterministic sensitivity analysis (DSA)

The primary objective of conducting a deterministic sensitivity analysis (DSA) is to examine the direction and magnitude of any possible changes in the study result over a reasonable fixed range, such as range, standard deviation, and 95% confidence interval of the model parameters. The simplest approach is to use a one-way SA, which determines the sensitivity of the result by examining variation between parameters, one by one(7). Tornado diagrams (Fig. 1) are a very useful tool for illustrating the degree of the result sensitivity with respect to uncertainty in each parameter.

In Fig. 1, the vertical line represents the point estimate (606,000 Baht/quality-adjusted life year, QALY) of the incremental cost-effectiveness ratio (ICER) obtained from a reference (or base case) analysis when adding a monoclonal antibody (rituximab) to the conventional chemotherapy (CHOP) for non-Hodgkin lymphoma(8). The length of each horizontal bar reflects the extent to which the study result would vary in accordance with a change in the model parameter on the vertical axis, given all other things being constant (as in the reference case). The most influential parameter stays at the top most (i.e. relative efficacy of rituximab on clinical response), whereas the least influential...
Threshold analysis

Threshold analysis is an approach that is used to determine the threshold value of a given parameter that would result in one treatment option being more cost-effective than the other. The analysis result can be described using a narrative text, for example ‘The ICER of …(option A)… is less than … Baht/QALY as long as …(parameter X)… is higher than …’; or ‘… (option A)… is more cost-effective than …(option B)… as long as …(parameter X)… is higher than …’.

It is not uncommon to find that a medicine’s unit price (or acquisition cost) can significantly influence the sensitivity of the analysis result. The cost-effectiveness of the treatment of interest can be analysed against the willingness to pay (WTP) threshold of societies. In this paper, the threshold of 120,000 Baht/QALY, the NLEM-recommended level at the time the paper was written, is used. In late 2013, the NLEM committee recommended raising the GNI to 160,000 Baht/QALY. The Thai Office of the National Economic and Social Development Board reported that the average national income was 153,952 Baht in 2010, according to data garnered using chain volume measures. In July 2011, Thailand was reclassified as an upper, middle-income country, as the per capita GNI, calculated according to the World Bank’s Atlas method, increased to USD 4,210. Therefore, in cases where the reference case ICER is above the national threshold, a threshold analysis is recommended to determine the reduced unit price of medicines.

A multi-way SA is undertaken when two or more parameters need to be taken into consideration at the same time. A graphical approach is very useful when presenting the results of a multi-way SA. Fig. 2 depicts the threshold plot using a two-way SA with three treatment options, A, B, and C. Contour lines represent the combined threshold values of two model parameters (horizontal axis for X and vertical axis for Y) that would clarify which one of the three options was most cost-effective at a certain WTP level.

Probabilistic sensitivity analysis (PSA)

Probabilistic sensitivity analysis (PSA) is increasingly becoming regarded as an indispensable tool in models where parameters are derived from research findings and individual databases, which are prone to uncertainty. An analysis of the economic evaluation model results can be repeated using random sampling of the model parameters based on appropriate assumptions of the data distribution.

Cost-effectiveness (CE) plane

Results from PSAs are usually presented initially using a CE plane, a two-dimensional diagram with four quadrants that gives an estimate of incremental cost (vertical axis) and incremental effectiveness (horizontal axis). Results of the PSAs are plotted on the diagram so that incremental cost and incremental effectiveness of the treatment of interest can be compared with a comparator, which is set at the 0-0 coordinate (origin). The slope of the straight line that connects the comparator at the origin and the treatment in quadrant 1 gives the ICER. This ICER value should

---

**Figure 1** Sensitivity of ICER for rituximab in non-Hodgkin lymphoma with respect to variations in drug efficacy, price and utility.

**Figure 2** Two-way threshold plot for three treatment options (A, B and C) Source: adapted from.
not be presented as valid if its value is less than zero, because while numerically valid, the implications for drawing conclusions from a negative ICER are likely to be vague or misleading. When results from the treatment of interest are located in either quadrant 2 (i.e. more effective and less costly) or quadrant 4 (less effective and more costly), this in itself is a clear indication of the treatment of interest’s superiority or inferiority.\(^\text{(17,18)}\)

Fig. 3 shows the results of an economic evaluation of gefitinib as compared with docetaxel in treating non-small cell lung cancer. The analysis comprised 1,000 repetitions of PSA, and the scatter of dots around the point estimate of the reference case reflect parameter uncertainty.

The vertical and horizontal dash lines in Fig. 3 that pass through the origin indicate that the total effectiveness (measured in QALYs) and total cost (measured in thousands of baht) of gefitinib are the same as those of docetaxel, the comparator. The result representing a reference case in quadrant 1 indicates that gefitinib is more effective and is more costly than docetaxel by approximately 0.06 QALY and 88.5 thousand baht, respectively, yielding an ICER of approximately 1.5 million baht/QALY.\(^\text{(18)}\)

The PSA that resulted in 1,000 dots in quadrants 1 and 2 suggests that gefitinib is always more effective than docetaxel. However, some PSA repetitions show gefitinib to be less costly (falling in quadrant 2) and some show it to be more costly (falling in quadrant 1) than docetaxel.

When the WTP threshold is greater than zero, the proportion of the dots below the level that indicates that gefitinib is cost-saving increases. For a given WTP threshold, the probability that a treatment is found to be cost-saving can be calculated using a net benefit calculation that takes into account both net monetary benefit\(^\text{(15,20)}\) and net health benefit\(^\text{(21)}\).

**Acceptability curve**

The cost-effectiveness acceptability curve represents how change in the probability of a treatment being cost-effective relates to change in the WTP threshold (Fig. 4)\(^\text{(16,22)}\).

The acceptability curve in Fig. 4 shows that there is a 40% probability that gefitinib is more effective and less costly than (or economically dominant over) docetaxel. The probability of gefitinib being found to be cost-effective increases gradually when the WTP increases. For instance, at WTP thresholds of 100,000; 300,000; 500,000, and 1,000,000 baht/QALY, the probability of gefitinib being found to be cost effective is 42%, 45%, 48% and 55% respectively. The fact that the acceptability curve for erlotinib is lower than that for gefitinib indicates that erlotinib is less cost-effective than gefitinib at the same WTP threshold.

**Expected value of information**

Where model parameters are uncertain, immediate decisions about whether a treatment should be regarded as optimal should be delayed while an assessment of the viability of further research to clarify the parameters is conducted. Uncertain parameters may lead to incorrect findings, which may contribute to a consequential loss. This kind of opportunity cost can
be calculated in terms of expected value of information\(^{(23)}\), an approach that is widely used in decision analyses\(^{(24)}\). An overall loss resulting from a wrong decision due to all-parameter uncertainty is called the “expected value of perfect information” (EVPI). This is equal to the difference between the best payoff (namely, expected value with perfect information or EV\(_{\text{wPI}}\)) and the expected value (EV) of making a decision when all the parameters are uncertain. Table 1 presents a calculation for the EVPI of three alternatives (A, B, and C), which result in outcomes X, Y, and Z.

A decision analysis conducted on the uncertainty of outcomes X, Y, and Z shows that option A is optimal, because it yields the maximum EV of 68,000 baht (EV\(_{\text{A}}\) = 200,000x0.3 + 40,000x0.5 - 60,000x0.2). Option B would yield a relatively lower EV of 42,000 baht (EV\(_{\text{B}}\) = 70,000x0.3 + 50,000x0.5 - 20,000x0.2), while Option C (for example, doing nothing) would yield the EV of 0 (zero).

If the probability for an outcome X, Y, and Z of 0.3, 0.5, and 0.2, respectively was known exactly, the maximum payoff among the three options for each outcome, which is equal to 200,000; 50,000, and 0 baht, respectively would yield the best payoff under this perfect information (EV\(_{\text{wPI}}\)) of 85,000 baht (EV\(_{\text{wPI}}\) = 200,000x0.3 + 50,000x0.5 + 0x0.2).

The EVPI is thus an opportunity cost incurred as a result of making the wrong decision under the current situation of data uncertainty, which is equal to the difference between the EV\(_{\text{wPI}}\) and the EV\(_{\text{A}}\) (in this case, 85,000 - 68,000 = 17,000 baht). The EVPI suggests that the decision should be delayed until further study results can be obtained. If in this case, the cost of conducting further research is higher than the EVPI of 17,000 baht, option A should be chosen immediately.

Because EVPI is derived using a net benefit approach, it should be presented in the same way as an acceptability curve. Fig. 5 illustrates the relationship between the EVPI and various degrees of the WTP threshold.

As Fig. 5 shows, if societies place a monetary value of 40,000 baht on an additional gain of one QALY, the potential loss resulting from a wrong decision under the current parameter uncertainty would be 800,000 baht, the maximum. In this case, the decision should be delayed until further research can be conducted, as long as that research costs less than 800,000 baht.

If however, the WTP threshold per QALY is less than 30,000 baht or more than 60,000 baht (as is the case in the current threshold suggested by the NLEM subcommittee), the opportunity cost of making the wrong decision will be very low.

The potential loss due to uncertainty resulting from any given set of parameters can be expressed in terms of expected value of partial perfect information’ (EV\(_{\text{PPPI}}\)). In practice, these parameters should be analysed as one common set if they are correlated\(^{(25,26)}\). If the EV\(_{\text{PPPI}}\) is found to be very high, further research on that particular set of parameters should be conducted.

**Guidelines for Health Technology Assessment in Thailand (second edition): Recommendations for handling uncertainty in economic evaluations**

In any economic evaluation that is based on a decision analysis model, SA should be reported alongside the reference case analysis. The simplest method for this is DSA, which can be part of either a one-way or multi-way SA. While a one-way SA, presented in the form of a tornado diagram, identifies each parameter that influences the analysis result, a threshold plot from the multi-way SA will help identify an appropriate decision threshold, where there are high

<table>
<thead>
<tr>
<th>Possible outcomes</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Payoff (in baht)</td>
<td>Option A: 200,000</td>
<td>40,000</td>
<td>-60,000</td>
</tr>
<tr>
<td></td>
<td>Option B: 70,000</td>
<td>50,000</td>
<td>-20,000</td>
</tr>
<tr>
<td></td>
<td>Option C: 0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 5 EVPI (in baht) and WTP threshold (in baht/QALY)\(^{(3)}\).
levels of uncertainty in the parameters. When uncertainty arises from several parameters simultaneously, PSA should be conducted. At any given WTP threshold, the probability of a cost-effective treatment being cost-effective can be detected using an acceptability curve. To delay a decision under the uncertainty while pursuing further research to obtain better data (where economically appropriate), an estimation of the expected value of information is helpful.

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 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
19. Limwattananon S, Sukprasert U, Waleekhachonlert


Budget Impact Analysis

Pattara Leelahavarong BSc, MSc*

* Health Intervention and Technology Assessment Program, Ministry of Public Health, Nonthaburi, Thailand

A budget impact analysis (BIA) is used to assess whether the adoption of a new health technology is affordable, given the resource and budget constraints of the context. Increasingly, BIAs are coming to be viewed as an important-if not essential-part of health technology assessment (HTA). BIA data is often examined in conjunction with cost-effectiveness analysis (CEA) data to help inform decision makers when developing reimbursement policies within the resource constraints of their health care system. This article presents a review of existing BIA guidelines from around the world and makes some initial recommendations for the development of Thai BIA guidelines, as part of the newly-developed Economic Evaluation guidelines for Thailand. Initial recommendations include guidelines on appropriate analytic framework design, study design, perspective, scenarios for comparison, target population, costing and resource use, uncertainty analysis, and discounting.

Keywords: Budget impact analysis, Economic evaluation, Thailand

A budget impact analysis (BIA) is used to assess whether the adoption of a new health technology is affordable, given the resource and budget constraints of the context. Increasingly, BIAs are coming to be viewed as an important-if not essential-part of health technology assessment (HTA). BIA data are often examined in conjunction with cost-effectiveness analysis (CEA) data to help inform decision makers when developing reimbursement policies within the resource constraints of their health care system. BIAs serve three main functions: 1) to estimate the financial consequences for a specified population of implementing a new health intervention or technology, 2) to provide data on the affordability of new health-care technologies at a given price for a specified population, prior to reimbursement, and 3) to serve as a budget or service planning tool that policy decision-makers can use to inform their allocation of resources once reimbursement of a given technology has been confirmed. Most HTAs will include both an economic evaluation and a BIA, the results of which should comprise the same data set and should be analyzed together as complementary findings. However, the two tools do differ in some key ways (Table 1), and their findings may sometimes conflict (for instance, where the economic evaluation indicates a technology does offer value for money but the BIA shows high budget impact). There is, unfortunately, no current scientific guidance on how to resolve these kinds of conflict.

Budget impact analytic framework

Fig. 1 shows an example BIA framework, which is adapted from the BIA framework developed by the International Society for Pharmaco-economics and Outcome Research (ISPOR) Task Force on Good Research Practices. The framework enables a comparison to be made between the treatment and condition status quo and the new situation that would result from the adoption of a new intervention. The BIA examines the impact of this adoption on the healthcare system by assessing its impact on a number of key factors (i.e. disease incidence, diagnosis and treatment, resource used, and costs). Finally, the total costs of each scenario are calculated and compared so that the budget impact of the adoption of the new technology can be estimated.

Six key inputs are required to construct the BIA modelling framework: 1) the size and characteristics of the affected population, 2) the current intervention mix, 3) the cost of the current intervention mix, 4) the proposed intervention mix, which will include the intervention under consideration, 5) the cost of the new intervention mix, which will include the intervention under consideration, and 6) the use and cost of other health treatment-related health-care services or
All relevant data should be researched, appraised, and presented according to the principles and methods of evidence-based medicine and systematic review.

**Review of published BIA guidelines**

The first analytic framework for BIA was published in 1998(7). Since then, a number of country-specific guidelines have specifically recommended that a budget impact analysis be included in any health technology economic evaluation where the findings will be used to inform national or local formulary approval or reimbursement decisions(7-12). To date, most BIA guidance has aimed at providing recommendations to ensure that policy-makers and those responsible for health care insurance budgets are provided with standardized, reliable, and good quality information. Until now, specific guidance on appropriate BIA methodologies, implementation, and good practice have largely been lacking. This article reviews existing methodological guidelines on a series of issues that should be considered by all who are involved in BIA development(1-4,6,8,13,14).

**Study design**

All existing guidelines recommend that a BIA should use the same input parameters, economic

---

**Table 1. Comparison of economic evaluation and budget impact analysis**(1,2,4)

<table>
<thead>
<tr>
<th>Detail</th>
<th>Budget impact analysis</th>
<th>Economic evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Underlying concept</td>
<td>- Affordability</td>
<td>- Value for money</td>
</tr>
<tr>
<td>Objective</td>
<td>- Financial impact of introducing a technology</td>
<td>- Economic efficiency of alternative technologies</td>
</tr>
<tr>
<td>Study timeframe</td>
<td>- As manager of the convenience (usually 1-5 years)</td>
<td>- Preferably lifetime</td>
</tr>
<tr>
<td>Health outcomes</td>
<td>- Excluded</td>
<td>- Included (e.g. quality-adjusted life years)</td>
</tr>
<tr>
<td>Perspective</td>
<td>- Budget holder/Manager</td>
<td>- Society/Third payers/other</td>
</tr>
<tr>
<td>Comparison</td>
<td>- Scenarios in which they can design the degree of incorporation of the new technology in population with a mixture of utilization</td>
<td>- Specific technologies: a new technology will be used throughout cohort intervention</td>
</tr>
<tr>
<td>Study population</td>
<td>- Open cohort: individuals can be included or excluded along time, considering rate incorporation of technology, incidence of disease indications and treatment effect of the new treatment on survival</td>
<td>- Close cohort: cohort of individuals defined a priori</td>
</tr>
<tr>
<td>Discounting</td>
<td>- Not recommended</td>
<td>- Highly recommended</td>
</tr>
<tr>
<td>Presenting result</td>
<td>- Total and incremental annual costs</td>
<td>- Incremental cost per unit of health outcome achieved</td>
</tr>
<tr>
<td>Generalisation of results</td>
<td>- Inadequate: budget impact studies are designed to specific circumstances</td>
<td>- Possible, with limitations</td>
</tr>
</tbody>
</table>

**Fig. 1** BIA framework. Adapted from International Society for Pharmacoeconomics and Outcome Research (ISPOR) Task Force on Good Research Practices(1).
models, and model assumptions as the economic evaluation, so that the results of both analyses can be examined together. Moreover, there is widespread agreement that all BIAs should also take into account the type of health condition (i.e. whether it is chronic or acute) before deciding upon an appropriate analytical approach (i.e. whether it is prevalence-based, incidence-based, or both) and type of intervention (i.e. whether it is preventive, curative, palliative, one-time, ongoing, or periodic). Decision modeling (e.g. Markov models and decision trees) are recommended instruments for use in BIAs, as they offer high levels of transparency and allow the model calculation formulae, model parameters, and findings of the analysis to easily be reviewed by other researchers.

**Perspective**

The primary objective of a BIA is to assess the affordability of incorporating a new technology within the existing health care insurance context. As a result, the budget holder perspective (whether within a national healthcare system, managed care organization, social insurance institution, or hospital context) is the perspective recommended by almost all BIA guidelines.

**Scenarios to be compared**

There is widespread agreement among existing guidelines that all scenarios that the BIA compares should be based on the reimbursement packages and mixed treatment interventions that would be implemented in reality for the target population. When making budgetary comparisons in BIAs, the new technology should only be analysed within the appropriate mixed treatment scenario; that is, the budget should be assessed in terms of total intervention budget, rather than an examination of the technology in isolation. Thus, the analysis should consider how the current mix of interventions is likely to change when the new intervention is made available\(^1\). This kind of analysis is different to that adopted by economic evaluations, which compare specific technologies on an individual basis, rather than examining how their adoption might change the current mixed treatment scenario(s). The existing guidelines make the following recommendations on the methods and data sources that should be used to generate these mixed treatment scenarios\(^1,2,4,6,8\):

1) Current technologies mixed treatment scenario: In many treatment programs, a number of treatment regimens are prescribed for patients with the same condition. As such, the proportion of patients undergoing each treatment needs to be estimated, and the resource use and costing values adjusted accordingly, before inclusion in the BIA. In addition, additional costs related to current treatments, such as those associated with managing side effects, related administrative costs, and costs associated with related procedures, should be calculated and included in the BIA. Consideration might also be made of any price discounts offered by the pharmaceutical reimbursement package and patient contribution charge.

**Recommended data source:** National reimbursement databases/health insurance databases

2) New technologies mixed treatment scenario:

The precise mix of the new technology’s treatment mix will depend on the rate of uptake of the new technology as well as the extent to which it will either replace or complement current technologies\(^1\). Moreover, the rate of uptake is likely to change over time, as physicians and patients become familiar with new technology.

**Recommended data source:** Producer estimates of market share

Where producer estimates of market share are not available, an extrapolated estimate of market share based on product diffusion data from either the same technology in a different setting or a similar technology in the budget-holder’s setting could be used.

**Target population**

Deciding on an appropriate target population size is a very important part of any BIA, as this can significantly affect the results of analysis. A number of data sources can be used to inform the decision, including epidemiology of disease data, the proportion of patients who are covered by health insurance, and the accessibility of health care services (usually given as a percentage of the total number of patients who are covered by health insurance). These factors should all be taken into account in the comparative analysis, with appropriate estimates made for the new technology treatment mix scenario. Some BIA guidelines recommend that no off-label use should be included in the dataset\(^1\), while other guidelines recommended they should be included, even though off-label use is not included in reimbursement packages\(^6\). Other guidelines suggest instead that off-label use should be included in the uncertainty analysis\(^3\). Table 2 shows the parameters and data sources that can be used to estimate target population size in a BIA. It also illustrates how the size can change over time, and how this too should be included, through open cohort analysis\(^13\).
Budget impact analyses should be presented within time horizons that are most relevant to the budget holder(1). BIA are typically concerned with costs over short time horizons (e.g. 1-5 years)(2-4,14); however, the general rule is that the time horizon should be able to capture the period within which meaningful differences between the costs and outcomes of competing technologies become apparent. This period will vary according to the conditions under which the intervention is to be introduced, and sometimes according to the predicted impact of the new intervention(1). However, generally, it will be longer than the current budget period because of the costs and benefits that accrue over time. In any case, results that can be disaggregated should be available over time within a period that is deemed appropriate by the budget holder(1,8).

**Table 2. Parameters and data sources that can be used to estimate target population size in a BIA**(9,13)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Data sources (ranked by the level of valid evidence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence and incidence (Total number of patients)</td>
<td>1. published country-specific literature, or 2. international epidemiology data may also be used, where validated by a Delphi panel of national clinical and epidemiology experts</td>
</tr>
<tr>
<td>Proportion of patients eligible and accessible for treatment with the current technology</td>
<td>1. relevant national registry databases, or 2. published literature, e.g. a previous BIA for medications with similar indications, or 3. estimated from Delphi panel of national clinical experts</td>
</tr>
<tr>
<td>Proportion of patients eligible for treatment with the new technology</td>
<td>1. published country-specific literature, e.g. a previous BIA for medications with similar indication, or 2. registered patients for the new technology, but it might be restricted prescription by reimbursement authorities, or 3. estimated from Delphi panel of national clinical experts</td>
</tr>
<tr>
<td>Proportion of eligible patients actually treated with the new technology</td>
<td>1. estimated from Delphi panel of national clinical experts</td>
</tr>
<tr>
<td>Other input parameters:</td>
<td>1. estimated from Delphi panel of national clinical experts</td>
</tr>
<tr>
<td>1) the annual growth in utilization of the new technology over the time horizon of the BIA</td>
<td>2) the changes in treatment patterns or guidelines and off-label used</td>
</tr>
</tbody>
</table>
| 3) the treatment sequencing | **Costing and resource utilization**

When BIA guidelines recommend adopting the budget holder’s perspective, the model should take into account the resources used and the costs incurred, including all direct medical care costs and all other costs that exert an impact on the budget or health care system. The charge levied for the new technology should be based on its value according to the benefit package—not its market price. The ISPOR Task Force’s guidelines(1) suggest that impacts on areas outside the health-care system—such as those incurred by loss of productivity—should not be included in a BIA, as these are not generally relevant to the budget holder. However, this may not apply when budget impact analyses are intended to inform the decision making of employers or private health insurers, nor in contexts where health-care systems rely on tax payments where lost production due to morbidity could have important implications for healthcare funding.

In general, therefore, all costs that might result from the introduction of the new technology within the BIA time horizon, including health outcomes and side effects, should be included, and the resource use profile should reflect the actual usage and the way the budget holder values these resources(1). Thus, the expenditure calculation will include all of the costs that are expected to result from adoption of the intervention (variable costs in the short-run and fixed and variable costs in the long-run). The published guidelines do not agree on whether or not future costs should be included for other health conditions that might be incurred in patients who survived as a result of the new intervention. The choice whether to include or exclude
future unrelated costs will depend on the payer requirement and perspectives.

Uncertainty analysis

BIAs include a considerable level of uncertainty. Therefore, a single “best estimate” is not a sufficient outcome. Instead, the BIA should compute a range of results that reflect the plausible range of circumstances the budget holder will face. It is useful to consider both the most optimistic and the most pessimistic scenario. The ranges that are presented must be based on realistic scenarios regarding various inputs and assumptions, such as the size of the target population, the different uptake rates for the new technology, the costs of the new technology, and other assumptions for which data were not available.

The identified plausible range of parameters and assumptions should be developed collaboratively with the decision makers, because they are best placed to make many of the key assumptions and to supply data for the ranges of input parameter values. In some cases, it may also be advisable for the decision-makers to review the BIA model to assess the scenarios therein and undertake their own sensitivity analysis. It should be understood by the decision-maker that some analyses might be sensitive.

Although a probabilistic sensitivity analysis (PSA) is required in any economic evaluation, the role of PSA within BIA has been questioned because of the accountability that PSAs requires. Moreover, while PSAs require the use of estimated variance data, input data used in BIAs often comes from panel expert discussion data, from which it is notoriously difficult to generate estimated variance. Most guidelines suggest that a deterministic sensitivity analysis should be conducted to identify the range of the budget impact.

Discounting

The published BIA guidelines recommend that discounting should not be factored into BIAs because discounted costs do not reflect the actual budget in given year.

Guidelines for Health Technology Assessment in Thailand (second edition): Recommendations for budget impact analysis

BIAs should be regarded as economic evaluation tools that policy decision can use in the following circumstances: 1) when a new technology seems to be cost-effective at either its current price or its submitted price used in the reimbursement package. In this case, the BIA should use the price that is most cost-effective; 2) where the new technology does not seem to be cost-effective, a threshold analysis should be conducted to assess whether the reduced price would make it value for money. This price could be used for negotiation processes, in which case the BIA should then use the negotiated price; or 3) when the target patient group of the technology under assessment is small (e.g. in the case for treatments for rare diseases), the treatment may not seem to be cost-effective. However, a BIA should be employed to inform policy decision-makers who might be interested in developing a reimbursement package for vulnerable patients.

Study design

BIAs should be conducted in conjunction with the economic evaluation; they should use the same input parameters, economic models, and model assumptions.

Perspective

All BIAs should be conducted from the budget holder’s perspective (whether within a national healthcare system, managed care organization, social insurance institution, or hospital context), given that that the new technology will impact upon their budget.

Scenarios to be compared

BIA scenarios should be based on existing reimbursement packages and should use sets of mixed treatment interventions (with and without the new technology) to see how the current treatment mix and the proposed new treatment mix would affect the target population. Thus, the resource and valuation measurement for each treatment that is included in the current intervention mix needs to be examined proportionately, using data obtained from the national reimbursement database/health insurance database. To estimate the new technology mix scenario, the rate of uptake should be forecasted from the producer’s estimates of market share or extrapolated from previous product diffusion data. Moreover, the rate of uptake is likely to change over time, as physicians and patients become familiar with new technology, so a Delphi panel of national clinical experts might be used to predict how the rate at which this will change.

Target population

A number of data sources can be used to
inform the decision, including epidemiology of disease data, the proportion of patients who are covered by health insurance, and the accessibility of health care services (usually given as a percentage of total patients who are covered by health insurance). These factors should all be taken into account in the comparative analysis, with appropriate estimates made for the new technology treatment mix scenario and within a suitable time horizon. As target population size can change over time, acknowledgement of this too should be included in the BIA, through open cohort analysis. Given that many interventions can be used to treat a number of conditions, it is important that any uses of the intervention that are not recommended in the reimbursement package should not be taken into account in the BIA. The calculation of target population should be clearly reported.

Time horizon

The time horizon used in the BIA should be that which is deemed most relevant to the budget holder. BIAs are typically concerned with costs over a short time horizon (e.g. 1-5 years).

Cost and resource utilisation

The costs and resource that are used in the BIA should be based on the budget holder’s perspective, and should thus include all direct medical care costs and all other costs that might impact the budget or health care system within the specified time horizon (e.g. 1-5 year-period). The charge levied for the new technology should be based on its value according to the benefit package—not its market price. The analysis should also include other costs that are related to the new intervention package, such as those resulting from side effects. Future costs associated with other health conditions that might be incurred due to patients surviving as a result of the new intervention should generally be excluded. However, the choice whether to include or exclude future unrelated costs will depend on the payer requirement and perspective.

Uncertainty analysis

A deterministic sensitivity analysis should be conducted to reveal the plausible range of budget impact, including both the most optimistic and most pessimistic scenario, as indicated by the Tornado diagram. The use of a PSA to examine the uncertainty of the CEA results may well be unnecessary in a BIA. The ranges that are presented must be based on realistic scenarios regarding various inputs and assumptions, such as the size of the target population, the different uptake rates for the new technology, and the costs of the new technology.

Discounting

It is not recommended that discounted costs be used in the BIA.

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

5. Orlewska E, Gulacsii L. Budget-impact analyses: a critical review of published studies. Pharmaco-


Presentation of Economic Evaluation Results

Usa Chaikledkaew BSc(Pharm), MA, 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

The first HTA guidelines for Thailand included a chapter outlining a set of guidelines on how best to report the findings of health economic evaluations, based on a review of best practice and existing guidelines on the presentation of economic evaluation results from around the world. In this second edition of HTA guidelines for Thailand, the recommendations build on the first edition by using a case study to illustrate how the guidelines can be applied in a real research context. The guidelines propose that all reporting include ten key elements: defining the scope of the study, selection of comparator(s), defining the type of economic evaluation, measurement of costs, measurement of clinical effects, handling time in economic evaluation studies, handling uncertainty and sensitivity analysis, presentation of the results, discussion of the results, and disclosure of funding and authors conflict of interest.

Keywords: Result presentation, Economic evaluation, Thailand

The first HTA guidelines for Thailand included a chapter outlining a set of guidelines on how best to report the findings of health economic evaluations, based on a review of best practice and existing guidelines on the presentation of economic evaluation results from around the world. In this second edition of HTA guidelines for Thailand, the recommendations build on the first edition by using a case study to illustrate how the guidelines can be applied in a real research context. The case study that was used examined the cost-utility and budget impact of allogeneic hematopoietic stem cell transplantation for severe thalassemic patients in Thailand. For detailed information on the theory of reporting for economic evaluations, and a full version of the recommendations, please refer to the first edition of the guidelines. A summary of the guidelines are given below, together with examples from the case study.

Concepts and Principles: Presentation of economic evaluation results

Defining the scope of the study

All introduction sections of study reports should include a summary of the background of the present study, study rationale, and the economic and a summary of the clinical importance of the study (prevalence, incidence, mortality rate, etc.). A summary of the economic burden of the intervention should also be provided, alongside a detailed description of the study design, and a description of the program or intervention under consideration. If a research question is identified, it should be in a format that can be answered with the results of the study. The research question should help to define the objective of the study, which should also be addressed in the introduction. In addition, because they affect cost calculation, the scope and boundaries of the research should be defined, including those related to the population, the type of effects or outcomes analysed, the time horizon, and the perspective adopted, the author gives a brief summary of these aspects in the example analysis. A summary of how the scope of the study was presented in the case study is given below.

The source of the problems associated with the treatment of patients with severe thalassemia was clearly identified by describing the nature of the disease and the incidence of the condition; this helped to justify the significance of the present study. Although hematopoietic stem cell transplantation (HSCT) is well-known as the only cure for severe thalassemia, it is very expensive when compared to the standard treatment-blood transfusion and iron chelating therapy (BT-ICT). In Thailand, three health
insurance schemes offer healthcare coverage to approximately 100 percent of the population—the Social Security Scheme, the Civil Servant Medical Benefit Scheme, and the Universal Coverage Scheme. At the time the research of the case study was conducted, only the Social Security Scheme and the Civil Servant Medical Benefit Scheme provided HSCT coverage for severely thalassemic patients. To assess whether HSCT provision was cost-effective and thus should also be included in the Universal Coverage Scheme, the National Health Security Office (NHSO) requested that a study be conducted into the cost-effectiveness of HSCT. The study compared the cost-effectiveness of HSCT with that of BT-ICT, the standard treatment option. A societal perspective was adopted, and a budget impact analysis was conducted from a governmental perspective to assess whether HSCT was sufficiently cost-effective to be included in the benefit package of the Universal Coverage Scheme. A cost-utility analysis using a model-based approach was applied to estimate the cost and quality adjusted life years (QALYs) throughout a lifetime period.

Selection of comparator(s)

All study reports should include details of all comparators and an explanation of why they have been chosen. This information is intended to guide real clinical practice. A summary of how this was presented in the case study is given below.

The main treatment options that were available for patients with severe thalassemia, as outlined in clinical practice guidelines—BT-ICT and HSCT—were described in detail, and a summary of the advantages and disadvantages of each option was given.

Defining the type of economic evaluation

All study reports should report clearly on the type of economic evaluation method used (from the four main types of economic evaluation method: cost-minimization analysis, cost-benefit analysis, cost-effective analysis, and cost-utility analysis), along with an explanation of how the method is the most appropriate to address the research questions. A summary of the study design should also be given (i.e. whether an economic evaluation model or economic evaluation with clinical trials is used). Where a mathematical or simulation model is used, all assumptions should be detailed, and the method used should be specified (i.e. whether a decision tree model, state-transition/Markov model, or a probabilistic simulation model is used). A diagram of the event pathway of the model and the software used should also be presented. Where the Markov model is used, health states, cycle length, mechanisms for movement between states in simulation models, and any special features of the analysis should be explained. For studies where a model is used, any tests that have been conducted to demonstrate the accuracy of the programming and to establish the face validity of the model calculations should be described in brief. By providing details on how the tests relate to the performance of the model using extreme assumptions, the predictability of the model’s results is demonstrated. A summary of how the type of economic evaluation used in the case study was presented in the case study is given below.

Fig. 1 shows the schematic diagram of the Markov model used in the case study. The following five health states with different costs and QoL scores were defined for both related and unrelated HSCT patients: 1) the first year of HSCT (where patients had the highest costs and worst QoL), 2) the second year of HSCT (where patients had higher costs due to follow-up visits and immunosuppressive therapy), 3) years following successful HSCT (where QoL is approximately equal to that of the healthy population and costs were vastly reduced), 4) where HSCT has failed (resulting in a switch to BT-ICT), and 5) death. Two health states were defined for blood transfusion-dependent patients: 1) BT-ICT (characterised by low QoL, and the costs of ongoing care), 2) death. In the diagram, arrows represent possible transitions from one state to another and details are given that explain the health state transition. In the model, cycle length was defined as one year, and costs and health outcomes were estimated over a 99-year period, to cover the
maximum expected lifetime horizon. All assumptions were specified. Specifically, it was assumed that all severely thalassemic patients were treated with blood transfusions during the first year of life, that ICT was administered via subcutaneous infusion only, and that the probability of death in HSCT failure patients when switching to blood transfusions was similar to that in blood transfusion patients that did not undergo HSCT.

**Measurement of costs**

In the cost calculation section of study reports, the study perspective should be described, and all costs and data sources used should be detailed. For instance, cost data could be collected from electronic databases maintained by hospitals, from interviews with patients, or by referring to standard cost lists for health technology assessment. Reporting on the sources of cost data in this way helps readers assess the quality of the cost data used in the study. In addition, the types of costs (direct and/or indirect) should be stated. It is also suggested that the average cost per unit of each resource should be given along with range values, the number of units consumed, the year in which the costs are presented in the study, the type of currency used, and exchange rate used. This information allows readers to interpret the incremental cost-effectiveness ratio and compare it with the results of other studies. Adjustments for inflation—such as use of the medical component of the Consumer Price Index (CPI)—should be specified where applicable. It is important that report include information on whether cost or charge data have been used, and whether the ratios of cost to charge have been applied. A summary table presenting all cost data used and the source of data for each estimate should be presented, as in the example given in Table 1.

**Measurement of clinical effects**

A summary of the effectiveness estimates determined according to reference case values used in the analysis should be presented in a table as a convenient reference for readers. All clinical variables should be presented in accordance with the method of economic evaluation used. For instance, if a cost-utility analysis method is applied, utility values should be presented. All mean values of effectiveness parameters and ranges used in the uncertainty analysis and sources of data should be reported. This will help the reader garner an understanding of the source of the analysis’s effectiveness, which is an essential part of evaluating the quality of the analysis and hence the appropriate use of its results.

If program-specific primary data are utilised in the model, the report should include details of the general strategy used, the inclusion and exclusion criteria applied, and the important assumptions that were made. For example, to extrapolate survival beyond the end of the empirical data, a survival analysis may have been used. If this is the case, any relevant assumptions that were made should also be addressed in the report. If a survey is used, the response rate should be reported. In addition, any information on health states or utility that may have been collected previously by measuring health states directly within the study, or by asking experts to determine the health states, should be outlined. The instruments that have been used for measurement (e.g. the Health Utilities Index or the EuroQoL), a summary of the measurement tool, as well as the methods used to value outcomes (e.g. rating scale or time trade off) should be reported. In the analysis, if experts are required to provide input (e.g., probabilities, costs, preference weights, etc.), the basis for selecting the experts, the source of their expertise, the number of experts contributing, the reason for using expert judgment, and the process used to obtain their input should be clearly described.

**Handling time in economic evaluation studies**

When a study period is longer than one year, the costs and health effects must be discounted. In these cases, the study report should state whether both costs and health effects are discounted and give the discount rate, along with a justification of the choice. If no discounting is performed in the study, an explanation of why this is the case should be provided. A summary of how the handling of time in the case study was presented in the case study is given below.

Both costs and outcomes were discounted at a rate of 3%, as recommended by the first Thai HTA guidelines.

**Handling uncertainty and sensitivity analysis**

All study reports should include a description of the methods that were used to evaluate the effects of uncertainty in the analysis—i.e. whether a univariate sensitivity analysis (a one-way sensitivity analysis and threshold sensitivity analysis), a multivariate sensitivity analysis, or probabilistic sensitivity analysis was used. All important results should be given, along with the confidence interval of the cost-effectiveness ratio. The choice of variables, the ranges used in the sensitivity analysis (e.g. confidence interval or standard error) and
reasons why selecting these variables also should be reported. If a probabilistic simulation model is used, any tests of the assumptions made concerning the distributions of variables and their statistical independence should be included.

**Table 1. Input parameters used in the model**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Distribution</th>
<th>Mean</th>
<th>SE</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yearly discount rate (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costs (range)</td>
<td>Gamma</td>
<td>3.00 (0-6.00)</td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>Outcomes (range)</td>
<td>Gamma</td>
<td>3.00 (0-6.00)</td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>Transition probabilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BT-ICT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual probability of death at age 0-1</td>
<td>Beta</td>
<td>0.010</td>
<td></td>
<td>[23]</td>
</tr>
<tr>
<td>Annual probability of death at age 2-5</td>
<td>Beta</td>
<td>0.003</td>
<td></td>
<td>[23]</td>
</tr>
<tr>
<td>Annual probability of death at age 6-10</td>
<td>Beta</td>
<td>0.002</td>
<td></td>
<td>[23]</td>
</tr>
<tr>
<td>Annual probability of death at age 11-15</td>
<td>Beta</td>
<td>0.010</td>
<td></td>
<td>[23]</td>
</tr>
<tr>
<td>Annual probability of death at age 16-20</td>
<td>Beta</td>
<td>0.025</td>
<td></td>
<td>[23]</td>
</tr>
<tr>
<td>Annual probability of death at age 21-30</td>
<td>Beta</td>
<td>0.015</td>
<td></td>
<td>[22]</td>
</tr>
<tr>
<td>Annual probability of death at age 31 and over</td>
<td>Beta</td>
<td>0.345</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td>HSCT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parametric survival: death</td>
<td>Lognormal</td>
<td>-8.07</td>
<td>2.00</td>
<td>Cohort</td>
</tr>
<tr>
<td>Constant for baseline hazard</td>
<td>Lognormal</td>
<td>0.16</td>
<td>0.06</td>
<td>Cohort</td>
</tr>
<tr>
<td>Age coefficient for baseline hazard</td>
<td>Lognormal</td>
<td>-0.61</td>
<td>0.41</td>
<td>Cohort</td>
</tr>
<tr>
<td>Ancillary parameter in Weibull distribution</td>
<td>Lognormal</td>
<td>-7.18</td>
<td>1.55</td>
<td>Cohort</td>
</tr>
<tr>
<td>Type of HSCT coefficient for baseline hazard</td>
<td>Lognormal</td>
<td>2.60</td>
<td>1.08</td>
<td>Cohort</td>
</tr>
<tr>
<td>Ancillary parameter in Weibull distribution</td>
<td>Lognormal</td>
<td>-0.74</td>
<td>0.34</td>
<td>Cohort</td>
</tr>
<tr>
<td>Resource cost parameters (THB)</td>
<td>Gamma</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total direct medical cost of related HSCT in the 1st year</td>
<td>Gamma</td>
<td>491,985</td>
<td>50,288</td>
<td>Hospital database</td>
</tr>
<tr>
<td>Total direct medical cost of related HSCT in the 2nd year</td>
<td>Gamma</td>
<td>42,694</td>
<td>15,535</td>
<td>Hospital database</td>
</tr>
<tr>
<td>Total direct medical cost of related HSCT in the following years</td>
<td>Gamma</td>
<td>11,638</td>
<td>3,240</td>
<td>Hospital database</td>
</tr>
<tr>
<td>Total direct medical cost of unrelated HSCT at the 1st year</td>
<td>Gamma</td>
<td>735,839</td>
<td>183,560</td>
<td>Hospital database</td>
</tr>
<tr>
<td>Total direct medical cost of unrelated HSCT at the 2nd year</td>
<td>Gamma</td>
<td>45,840</td>
<td>20,094</td>
<td>Hospital database</td>
</tr>
<tr>
<td>Total direct medical cost of unrelated HSCT in the following years base</td>
<td>Gamma</td>
<td>6,385</td>
<td>1,037</td>
<td>Hospital database</td>
</tr>
<tr>
<td>Total direct medical cost of BT-ICT per year</td>
<td>Gamma</td>
<td>35,788</td>
<td>4,156</td>
<td>[4]</td>
</tr>
<tr>
<td>Total direct non-medical cost of HSCT at the 1st and 2nd year</td>
<td>Gamma</td>
<td>259,994</td>
<td>95,355</td>
<td>Survey</td>
</tr>
<tr>
<td>Total direct non-medical cost of BT-ICT and the following year of HSCT</td>
<td>Gamma</td>
<td>37,384</td>
<td>7,040</td>
<td>Survey</td>
</tr>
<tr>
<td>Total productivity loss of HSCT in the 1st and 2nd year</td>
<td>Gamma</td>
<td>77,468</td>
<td>70,464</td>
<td>Survey</td>
</tr>
<tr>
<td>Total productivity loss of BT-ICT and the following years of HSCT</td>
<td>Gamma</td>
<td>19,171</td>
<td>6,692</td>
<td>Survey</td>
</tr>
<tr>
<td>Utility parameters</td>
<td>Beta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility of BT-ICT patients</td>
<td>Beta</td>
<td>0.61</td>
<td>0.16</td>
<td>[24,25]</td>
</tr>
<tr>
<td>Utility of HSCT patients in first and second year</td>
<td>Beta</td>
<td>0.61</td>
<td>0.16</td>
<td>[24,25]</td>
</tr>
<tr>
<td>Utility of HSCT patients from third year on</td>
<td>Beta</td>
<td>0.93</td>
<td>0.05</td>
<td>[26]</td>
</tr>
</tbody>
</table>

BT-ICT = blood transfusion combined with subcutaneous iron chelating therapy; HSCT = hematopoietic stem cell transplantation; THB = Thai baht in 2008 value

**Presentation of the results**

**Presentation of incremental cost-effectiveness ratio (ICER) results**

Reference case results should be presented as a table of costs and effects for all the alternatives.
For each alternative, it is recommended that per capita of total costs, total effectiveness, incremental costs, incremental effectiveness, and incremental cost-effectiveness ratios (ICER) be provided in an accompanying table. Discounted results (using the discount rate at reference case) should be presented as the main results, while the undiscounted results may be put in the appendix. The following results should be presented:

1. Total cost per capita
2. Effectiveness per capita presented as both life years (LYs) and quality adjusted life years (QALYs). This helps readers understand the significance of extending life expectancy and improving quality of life as a result of each alternative.
3. Incremental cost per capita
4. Incremental effectiveness per capita and
5. Incremental cost-effectiveness ratio (ICER) per capita

The costs and incremental cost-effectiveness ratios should also be rounded up, either to a whole baht or to the nearest thousand, whichever is deemed most appropriate given the scale of the costs. The effectiveness data should also be rounded up where appropriate, as can be seen in the example in Table 2. Disaggregated results on costs, outcomes, and cost-effectiveness ratios should be presented so that the reader has an adequate understanding of the costs and effects of the intervention. For instance, total medical costs and total non-medical costs or QALYs, classified by disease severity should be disaggregated and included in the study report. The number of LYs saved and QALYs saved should be reported to help readers understand how the life-lengthening and quality-enhancing benefits of the intervention compare.

Moreover, the ICER results at both a population and an individual level should be reported. It is not recommended that the average or absolute ICERs for each alternative be reported, as this may lead to confusion and then misinterpretation of results(1). For instance, in the case study, ICER results were disaggregated according to the age of the patient (ranging from 1 to 28 years); this was because the patient’s age was found to have an impact on the success rate of HSCT, which in turn affects the cost-effectiveness of HSCT (Table 2). ICER results should be presented in THB, in terms of cost per unit of effectiveness according to the year of the cost calculation, for example 100,000 baht per QALY gained (2013 baht value).

Where the ICER results are negative, study reports should not frame the findings as negative or absolute values. Since negative values can imply two different meanings (i.e. higher cost and lower effectiveness or lower cost and higher effectiveness than other options), confusion may arise from the use of negative or absolute terminology. To avoid this, rather than reporting the ratios, the terms’ dominated’ should be used where the option has higher cost and lower effectiveness and ‘dominant’ where the option has lower cost and higher effectiveness compared to other options.

<table>
<thead>
<tr>
<th>Age (year)</th>
<th>Incremental cost per million THB</th>
<th>Incremental QALY gained</th>
<th>ICER of unrelated HSCT compared to BT-ICT THB per QALY gained*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.96</td>
<td>4.57</td>
<td>209,000</td>
</tr>
<tr>
<td>5</td>
<td>0.94</td>
<td>4.16</td>
<td>225,000</td>
</tr>
<tr>
<td>10</td>
<td>0.91</td>
<td>3.05</td>
<td>297,000</td>
</tr>
<tr>
<td>15</td>
<td>0.84</td>
<td>0.87</td>
<td>953,000</td>
</tr>
<tr>
<td>17</td>
<td>0.80</td>
<td>0.26</td>
<td>3,270,000</td>
</tr>
<tr>
<td>18</td>
<td>0.78</td>
<td>-0.01</td>
<td>Dominated**</td>
</tr>
<tr>
<td>19</td>
<td>0.73</td>
<td>-0.57</td>
<td>Dominated**</td>
</tr>
<tr>
<td>20</td>
<td>0.68</td>
<td>-1.12</td>
<td>Dominated**</td>
</tr>
<tr>
<td>25</td>
<td>0.59</td>
<td>-2.28</td>
<td>Dominated**</td>
</tr>
<tr>
<td>28</td>
<td>0.60</td>
<td>-2.22</td>
<td>Dominated**</td>
</tr>
</tbody>
</table>

Table 2. ICER of unrelated HSCT compared to BT-ICT, classified by patient age

ICER: incremental cost-effectiveness ratio; HSCT: hematopoietic stem cell transplantation; BT-ICT: blood transfusion combined with subcutaneous iron chelating therapy; THB: Thai baht (in 2008 value); and QALY: quality adjusted life year.

* ICERs are rounded up to nearest 1,000 THB.

** Negative ICER due to higher effectiveness and lower costs of BT-ICT compared with HSCT.
In most cases, graphical presentations of study results are recommended, as this can aid comprehension of the results. If the ICER results are presented graphically as a cost-effectiveness plane, the incremental costs (two consecutive interventions) should be displayed on the vertical axis and the incremental effectiveness (i.e., QALYs) should be on the horizontal axis, so that the slope of the line segment represents the incremental cost-effectiveness ratio.

**Presentation of uncertainty analysis results**

If a one-way sensitivity analysis method is performed, a tornado diagram showing the percentage change in the ICER attributable to the change of each individual parameter should be presented. The numbers at each end of the bars should indicate the most extreme values used in the sensitivity analysis, as shown in Fig. 2.

In addition, if a probabilistic sensitivity analysis is performed, cost-effectiveness acceptability curves, which present the relationship between the value of ceiling ratios (willingness to pay for a unit of outcomes) and the probability of favouring each treatment strategy, should also be presented in the study report, as shown in Fig. 3. These graphs demonstrate the probability of each intervention being cost-effective at different ceiling ratios, classified by age (year) at the start of treatment. (A) Patient aged 1 year, (B) Patient aged 10 years, (C) Patient aged 15 years, and (D) Patient aged 17 years. Dashed lines represent the thresholds for the adoption of health interventions in Thailand.

**Presentation of budget impact results**

For studies where a budget impact analysis is conducted, the study report should include information on the possible budget impact on total healthcare costs in both public and private sectors. The data should be given in a table that includes the year of calculation, expected total annual budget for each alternative, incremental budget per year, and expected total budget throughout the relevant period. The results of the budget impact analysis should also be rounded up to the nearest whole baht or the nearest million, depending on which is deemed most appropriate given the scale of the costs, as shown in Table 3.

**Discussion of the results**

The discussion section should provide an overview and interpretative summary of the results as well as a summary of any assumptions that were adopted. The impact of the findings on the results and the assumed impacts to the health system, health expenditures, and health equity should be discussed. A descriptive interpretation of the ICER results for the reference case should be given, and the results of the sensitivity analysis of key parameters should be discussed. If some parameters are suspected of causing biased results, the effects of those should be discussed.

The cost-effectiveness of an intervention can only be determined relative to other interventions. It is
difficult to make certain statements regarding the incremental cost-effectiveness ratio of an intervention by examining it in isolation. Whether an intervention should be implemented depends on the resources available, alternative uses of resources, and other constraints considered by decision-makers. As a result, researchers should be careful when stating that an intervention is ‘cost-effective’ or ‘not cost-effective’. Whether or not an intervention should be adopted will depend on multiple factors determined by the context. It is therefore not recommended to use only the cost-effectiveness criteria as the information for policy decision-making.

Where the results of a health economic evaluation may answer a specific policy question, the relevance of the study’s results should be clearly explained. Not all interventions should be evaluated only in terms of value for money. Therefore, widely-used alternatives should also be discussed in order to apply cost-effectiveness results in a broader context. A comparison of results from other economic evaluation studies of similar or related interventions should be included, and a discussion of the similarities and differences between the results of the studies should be clearly explained. To ensure results are comparable across studies, all currency rates and values should be converted to rates equivalent to those used in the author’s study, by applying the Consumer Price Index (CPI). If the year of analysis is not specified in the report, it is suggested that three years before the date of publication be assumed.

It is important that the present study report include acknowledgement that other factors, a side from cost effectiveness, are important when assessing a technology. For instance, it is important that the discussion take into account the potential budgetary impact for public and private healthcare expenditure if the technology is adopted. This discussion should assess the impact on annual budgeting and the cumulative impact over a relevant period. It is also important to highlight the possible savings or additional non-monetary resources that will be needed when the intervention is implemented. It may be important to discuss whether the introduction of the intervention will lead to increased or decreased need for related health care services. Furthermore, the discussion should also take into account equity or ethical considerations related to the introduction of the new intervention, for example potential impacts on access or utilisation of healthcare, reduced or

<table>
<thead>
<tr>
<th>Fiscal year</th>
<th>Estimated budget impact (million THB)</th>
<th>Incremental budget</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BT-ICT</td>
<td>Related HSCT</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>--------------</td>
</tr>
<tr>
<td>2008</td>
<td>7</td>
<td>98</td>
</tr>
<tr>
<td>2009</td>
<td>14</td>
<td>104</td>
</tr>
<tr>
<td>2010</td>
<td>20</td>
<td>103</td>
</tr>
<tr>
<td>2011</td>
<td>26</td>
<td>102</td>
</tr>
<tr>
<td>2012</td>
<td>32</td>
<td>101</td>
</tr>
<tr>
<td>2013</td>
<td>37</td>
<td>100</td>
</tr>
<tr>
<td>2014</td>
<td>42</td>
<td>99</td>
</tr>
<tr>
<td>2015</td>
<td>46</td>
<td>99</td>
</tr>
<tr>
<td>2016</td>
<td>50</td>
<td>98</td>
</tr>
<tr>
<td>2017</td>
<td>54</td>
<td>97</td>
</tr>
<tr>
<td>2018</td>
<td>58</td>
<td>96</td>
</tr>
<tr>
<td>2019</td>
<td>61</td>
<td>95</td>
</tr>
<tr>
<td>2020</td>
<td>64</td>
<td>93</td>
</tr>
<tr>
<td>2021</td>
<td>67</td>
<td>92</td>
</tr>
<tr>
<td>2022</td>
<td>69</td>
<td>91</td>
</tr>
<tr>
<td>Total</td>
<td>647</td>
<td>1,468</td>
</tr>
</tbody>
</table>

BT-ICT = blood transfusion combined with subcutaneous iron chelating therapy; HSCT = hematopoietic stem cell transplantation; THB = Thai baht (2008 value)
increased inequality in health status, and effects on disadvantaged social groups, should also be discussed.

Moreover, the limitations of the present study should be discussed to help interpret and generalize the results. All assumptions that have been made, whether based on expert opinions, theoretical models, or incomplete data, should also be stated as limitations. Often, the limitations result from nuances or complexity within the study results, which make the findings difficult to apply directly to policy decision-making.

Disclosure of funding and author’s conflict of interest

The present study should reveal the source of financial support of the present study to ensure transparency. The relationship between the authors and financial supporters and/or the authors’ potential conflict of interest with the funding sources should also be specified. In general, financial support may be stated in the acknowledgement section.

Guidelines for Health Technology Assessment in Thailand (second edition): Recommendations for presenting economic evaluation results

The Thai HTA guidelines recommend that the following ten key elements be included in any presentation of an economic evaluation study.

1) Defining the scope of the study
2) Selection of comparator(s)
3) Defining the type of economic evaluation
4) Measurement of costs
5) Measurement of clinical effects
6) Handling time in economic evaluation studies
7) Handling uncertainty and sensitivity analysis
8) Presentation of the results (i.e. ICER, uncertainty analysis and budget impact analysis)
9) Discussion of the results, limitations, impact to health system, expenditure and equity
10) Disclosure of funding and author’s conflict of interest

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 (HTAP) 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

Social and Ethical Analysis in Health Technology Assessment

Sripen Tantivess BSc, MPH, PhD*

* Health Intervention and Technology Assessment Program, Ministry of Public Health, Nonthaburi, Thailand

This paper presents a review of the domestic and international literature on the assessment of the social and ethical implications of health technologies. It gives an overview of the key concepts, principles, and approaches that should be taken into account when conducting a social and ethical analysis within health technology assessment (HTA). Although there is growing consensus among healthcare experts that the social and ethical ramifications of a given technology should be examined before its adoption, the demand for this kind of analysis among policy-makers around the world, including in Thailand, has so far been lacking. Currently, decision-makers mainly base technology adoption decisions using evidence on clinical effectiveness, value for money, and budget impact, while social and ethical aspects have been neglected. Despite the recognized importance of considering equity, justice, and social issues when making decisions regarding health resource allocation, the absence of internationally-accepted principles and methodologies, among other factors, hinders research in these areas. Given that developing internationally agreed standards takes time, it has been recommended that priority be given to defining processes that are justifiable, transparent, and contestable. A discussion of the current situation in Thailand concerning social and ethical analysis of health technologies is also presented.

Keywords: Social analysis, Ethical analysis, Health technology assessment

In most countries, policy-making decisions concerning allocation of resources to new healthcare technologies are made on the basis of evidence that assesses the clinical effectiveness, safety, and value for money of the technology under consideration. However, these are not the only issues that need to be taken into account when making decisions about the adoption of various technologies. Indeed, there is increasing evidence that policy-makers should also consider the potential negative social and ethical effects of adopting a given technology(1). Although social and ethical analyses have been a recommended part of health technology assessments (HTAs) since the 1970s(2), the number of studies on ethical analysis in HTA is still relatively small, compared with the total number of HTA publications. This might be owing to the absence of internationally-agreed principles, concepts, and methods(3,4), as well as inadequate understanding, knowledge, and skills among HTA researchers.

In Thailand, however, social and ethical reasons have occasionally been cited to justify the inclusion of health interventions in government-financed benefit packages, including the revision of the National List of Essential Medicines (NLEM)(5). In some instances, social and ethical reasons played an even more important role in coverage decisions than did the cost-effectiveness and fiscal impact of the technology under consideration. Nevertheless, no explicit guidance has yet been developed for how to conduct these kinds of analyses, nor for how to apply the findings to the policy development process (the first edition of HTA guidelines for Thailand did not include recommendations on social and ethical analysis).

To go some way to address this lack, and to begin the process of developing evidence-based and informed guidance for policy makers and researchers, this paper surveys the domestic and international literature concerning the evaluation of social and ethical implications of health technologies and outlines the main concepts, principles, and approaches. An initial discussion on how these relate to the current situation...
Why social and ethical assessment of health technology is important

There is widespread agreement among HTA experts that integrating social and ethical elements in HTAs is an important way to improve transparency and provide understanding on the context into which certain technologies will be introduced. Crucially, since the introduction and implementation of innovative technologies, will inevitably result in changes in their supply and demand, inequitable accessibility may also result. For this reason in particular, conducting HTAs without an ethical analysis is to ignore a crucial dimension of the technology under consideration. Although clinical benefits and cost-effectiveness are key concerns of policy-makers, their assessment often requires specific expertise, which is usually communicated in specialized language. In contrast, ethical analysis involves issues and language that is commonly understood, and the discussion of which encourages wider participation and improved transparency, resulting in well-accepted decisions and more effective implementation.

The implementation of new health technologies may also have social, legal, and political consequences, including those that may affect organisations. To ensure that these consequences are anticipated and addressed, experts recommend that HTA be expanded so that appropriate analysis in relevant disciplines be conducted to anticipate what, if any, consequences of this kind might arise from the introduction of a new technology. Guidelines published by the European Network for Health Technology Assessment (EUnetHTA) identify a number of key ways in which the introduction of new health technologies might affect patients, their family, and caregivers, in areas such as professional life, skills, and social position. While this does show an awareness of the social and ethical issues involved in introducing a new technology, this is still a largely isolated example; specific published guidance on social and ethical evaluation of health technologies remains limited.

Concepts and principles of analysis

Although research on social and ethical implications has been recognized as an important component of HTA for decades, there is, as yet, no common agreement on its principles and methodology. It is also unclear under what conditions a technology should be subject to this kind of analysis, who should be responsible for such analysis, and whether and to what extent key stakeholders and the general public should take part in the processes.

Ethics refers to philosophical study on the nature and grounds of moral judgments and standards and rules of conduct, both actual and practical. Put simplistically, ethical inquiries involve examining the differences between actions, behaviors, and ways of living, which are classified accordingly on a scale of good to bad. As coverage decisions determine access to and benefit from health technologies—which can both prolong life or cause adverse reactions—such policies should be regarded as ethical issues. The justification of whether the use of a selected technology is ‘right’ or ‘wrong’ depends on the social, cultural and political context within which the technology is to be implemented. For example, blood transfusion, abortion, and sex education are permitted in some countries, while unacceptable in other societies for religious or moral reasons.

An ethical analysis of health technologies should not rely on theories from one discipline alone; instead, a number of theories, all of which are accepted widely in the society under consideration, should be used. A review of the existing literature on ethical analyses of new technologies suggests that the most widely-introduced ethical principles are as follows: 1) biomedical ethics, which comprise respect for autonomy, i.e. the rights of people to acquire necessary information and make their own decisions to seek care and use certain technologies; 2) non-maleficence, i.e. the duty of health providers to avoid causing harm, suffer, injury, disability, or fatality intentionally; 3) beneficence, i.e. to enable people to have a ‘good life’, including harm reduction and prevention of negative consequences; 4) justice, which involves fair allocation of fundamental social burdens and benefits, i.e. to provide access to essential health services to all people, regardless of personal characteristics such as nationality, colour, and socioeconomic status. Biomedical ethics also relate to human dignity—a fundamental right of mankind.

According to EUnetHTA Guidelines, social analyses of health technologies should emphasise “patient-centred” principles, by examining the impact the health technology will have on individuals, including patients, caregivers and family members. This includes analyzing the extent to which doctors and healthcare providers provide patients with information about the technology, assessing the level of patient and caregiver understanding, and exploring the
feasibility of patient participation in decision making. HTA should also include assessment of how the intervention might affect the major areas of the life of the patient or caregiver, for instance, any impact it might have on work life, family life, leisure time, and religious and cultural activities.

The EUnetHTA guidelines also emphasize the point that every step of HTA includes social or ethical issues, including the selection of the technology; stakeholder participation; identification of social and ethical queries, and the selection of study design, method, and references. Every process in the HTA will have ethical ramifications and, as such, every stage in the assessment process should be cognisant of these issues. HTA experts agree that any assessment of the effectiveness and safety of a health technology should involve not only technical expertise, but also social and ethical consideration of the consequences of both adopting the technology and HTA process itself(6,13).

Assessment processes and approaches

As discussed earlier, EUnetHTA suggests that ethical implications should be considered at every stage of an HTA, from identification of the research questions to report writing and publication. They also suggest that the ethical analysis not be conducted in isolation but, instead, in tandem with all other assessments, as every HTA process, including clinical trials and economic evaluations, are value-laden(7). However, some experts have suggested that ethical assessment, as a unique type of investigation, should never be performed together assessments of the clinical, social, economic and legal implications of technology(14). Others argue that in absence of common agreement on the most appropriate methods for ethical analysis, emphasis should be placed on transparent, fair, reliable, and contestable processes for resource allocation, rather than detailed assessment guidelines.

Several academics have suggested that the general research methods usually deployed in HTAs (such as conducting primary research, transferring secondary data from existing studies and literature, and seeking expert advice(15)) are inappropriate for use in ethical analyses due to the philosophical nature of these kinds of investigations(16). Moreover, although international organizations such as the International Network of Agencies for Health Technology Assessment (INAHTA) and EUnetHTA have formulated guidelines for evaluating ethical implications of health technologies (in 2005 and 2008, respectively), these have been subject to criticism.

In practice, researchers propose sets of questions, developed on the back of ethical principles and theories. For example, Hofman(15) formulates 33 questions in 5 categories, comprising questions on moral issues, stakeholders, technology, assessment approaches, and technology assessment. In a similar vein, EUnetHTA recommends 14 questions on key ethical aspects of technology, autonomy, human dignity, beneficence and non-maleficence, justice and equity, and rights and legislation(7). For an evaluation of social implications, this HTA organisation network suggests 9 questions that focus on three areas-major life areas, individual impacts, and patient-healthcare provider communication. Table 1 presents selected questions from EUnetHTA’s Guidelines that refer to social and ethical issues.

Analysis of social and ethical implications of health technology in Thailand: Current situation and challenges

As mentioned previously, there are currently no national guidelines for conducting social and ethical analyses of health technologies in Thailand. Indeed, the demand for related evidence among policy-makers and stakeholders is lacking. Despite this, ethical issues have occasionally come to the fore, particularly in terms of inadequate access to high-cost technologies. Civil society organisations, patient groups, and healthcare professionals have begun to campaign to end what they see as inequitable access to various treatments, including medicines for antiretroviral treatment(18) and renal replacement therapy for end-stage renal disease(17). Certain groups have also advocated the government to expand the use of TRIPs flexibilities(18,19), and there have been protests against the expansion of intellectual property protection beyond TRIPs in the Thai-USA and Thai-EU Free-Trade Agreements.

Clearly, over the past few decades, social and ethical statements have played a significant role in Thailand’s health policy processes. Although no official mechanism has ever been in place to address such issues, a number of studies have generated useful, policy-relevant evidence on social and ethical issues of healthcare. Key studies include an equity analysis of how contributing to national health benefit schemes affects people in different income groups, a study that estimated the health and economic loss among poor households that resulted from inadequate access to renal replacement interventions(20), and a study focusing on the consequences of extending the pharmaceutical markets beyond the exclusivity of the
Table 1. Selected questions on the social and ethical aspects of health technologies (from EU net HTA’s Guidelines)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Social aspect</strong></td>
<td>Major life areas Which social areas does the use of the technology influence? Who are the important others that the use of the technology may affect in addition to the patient?</td>
</tr>
<tr>
<td>Individual</td>
<td>How do patients and important others react and act upon the technology?</td>
</tr>
<tr>
<td>Communication</td>
<td>What is patients’ and important others’ knowledge and understanding of the technology? How is the information regarding the use of the technology processed and exchanged?</td>
</tr>
<tr>
<td><strong>Ethical aspect</strong></td>
<td>Ethical aspect of technology Can the technology challenge religious, cultural, or moral convictions or beliefs of some groups or change current social arrangements?</td>
</tr>
<tr>
<td>Autonomy</td>
<td>Is the technology used for patients/people that are especially vulnerable?</td>
</tr>
<tr>
<td>Human dignity</td>
<td>Does the implementation or use of the technology affect human dignity?</td>
</tr>
<tr>
<td>Beneficence/non-maleficence</td>
<td>What are the benefits and harms for patients, and what is the balance between the benefits and harms when implementing and when not implementing the technology? Who will balance the risks and benefits in practice and how? Can the technology harm any other stakeholders? What are the potential benefits and harms for other stakeholders, what is the balance between them? Who will balance the risks and benefits in practice and how?</td>
</tr>
<tr>
<td>Justice and equity</td>
<td>What are the consequences of implementing/not implementing the technology on justice in the health care system?</td>
</tr>
<tr>
<td>Rights</td>
<td>Does the implementation or use of the technology affect the realisation of basic human rights?</td>
</tr>
<tr>
<td>Legislation</td>
<td>Is legislation and regulation to use the technology fair and adequate?</td>
</tr>
</tbody>
</table>

World Trade Organisation’s TRIPs agreement\(^{(21)}\). Nevertheless, while useful, these studies only cover the social and ethical implications of specific technologies and issues.

Social and ethical aspects are also addressed in a number of qualitative studies, including those conducted as part of an HTA. For example, one recent study examined the feasibility of stem cell transplantation for the treatment of severe thalassemia under Thailand’s Universal Health Coverage (UC) plan\(^{(22)}\). The findings suggest that, although inclusion of this technology in the UC benefit package may result in increased accessibility, the country’s capacity to provide stem cell transplantation is limited, and patients in low socioeconomic groups are less likely to comply with selection criteria for transplantation than those who are better off. This kind of inequity issue, which very important, is not captured in the two predominant forms of analysis favoured by policy-makers and is often not captured by most HTA.

The absence of national guidelines on conducting social and ethical analyses as part of HTAs may well result in opaque decision-making, and allow the assessment process to be influenced by policymaker bias and other unacceptable determinants, such as the preferences of manufacturers of technologies. In such cases, policy decisions might have negative consequences for patients, caregivers, and society, including unique table patient access and unfair treatment. The under privileged in society are likely to be disproportionately affected. Reviews of the current situation, arguments, and limitations concerning capacity for social and ethical assessment of health technologies indicate that Thailand needs to urgently address whether social and ethical analysis should be part of HTA. This depends, in part, on demands for related evidence among policy-makers and key stakeholders in the area of health priority setting and resource allocation, as well as related research capacity in the country.
Guidelines for Health Technology Assessment in Thailand (second edition): Recommendations for social and ethical analysis

HTA institutes in Thailand should adopt transparent, reliable, and contestable assessment processes, which allow the participation of all stakeholders. HTA research should be academically robust, with appropriate strategies to manage conflict of interest. In addition, researchers may consider whether the social and ethical questions and analytical approaches suggested in EUnetHTA Guidelines are appropriate for the Thai setting, and adopt elements where suitable. Moreover, wherever HTA methods may have social and ethical consequences, researchers should include a discussion of these issues in their reports.

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 (HTAP) 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

7. EUnetHTA. HTA Core Model for Medical and Surgical Interventions, 1.0R [Internet]. 2008 [cited 2012 Oct 4]. Available from: http://www.eunethta.eu/sites/5026.fedimbo.belgium.be/files/HTA%20Core%20Model%20for%20Medical%20and%20Surgical%20Interventions%201.0r.pdf
A Way Forward for the Evaluation of Health Technologies for Infectious Diseases

Pritaporn Kingkaew BPharm, MSc*

* Health Intervention and Technology Assessment Program (HITAP), Nonthaburi, Thailand

Recently, many researchers undertaking health technology assessment of screening or vaccination programs for infectious diseases have opted for dynamic transmission models for their analysis, rather than the typical static models (Markov and decision tree), as they are better at predicting the indirect effects of interventions, such as those that may affect disease transmission within the interested population or the ecology of the pathogen. Nevertheless, these models have not yet become part of the traditional tool box of health economists, due in part to the fact that the results are complex and difficult to analyze, requiring extensive computational skills. This paper aims to provide an overview of the concept of a dynamic transmission models and outline recommendations on how best to determine whether a dynamic approach is appropriate when evaluating health technologies and interventions for infectious diseases.

Keywords: Technology assessment, Cost-benefit analysis, Dynamic transmission model, Infectious diseases, Vaccines

Since the mid-1990s, the number of studies on the evaluation of health technologies and interventions related to infectious diseases has increased significantly. This rise in interest seems to have been the result of a number of changes that occurred in the last two decades, including the rising incidence and spread of certain infectious diseases, better understanding of the causes of various infectious diseases, and above all, the development of novel vaccines that can be used to prevent infection. One recent systematic review showed that, of the economic evaluation studies published between 1990 to 2009, the proportion of economic evaluations that examined infectious diseases rose from 8% to 13%, with most of these evaluating vaccines. Vaccination programs, especially universal vaccination programs, are widely regarded as worthwhile interventions, and their use in public health programs is often assumed to be cost-effective. However, whether a vaccination program is cost-effective will depend on a variety of factors, including the efficacy of the vaccine, acceptance rates, prices set, and society’s willingness-to-pay. Therefore, the value of such interventions will vary according to the specific context within which the decision-making and implementation is taking place; this can be seen in recent evaluations of vaccines for the Human immunodeficiency virus (HIV) and Human Papilloma virus (HPV), which showed marked differences in value depending on the context of the program.

Mathematical modelling can be used to simulate the progression of infectious diseases. In the public health arena, these models are usually used to evaluate public health policies by forecasting incidences of emerging diseases resulting from different policy scenarios, such as livestock culls, travel restrictions, school closures, or isolation and quarantine of infected persons/animals. Recently, many researchers undertaking health technology assessment of screening or vaccination programs for infectious diseases have opted for mathematical models (sometimes referred to as dynamic transmission models or dynamic models) for their analysis, rather than the typical static models (Markov and decision tree), as mathematical models are better at predicting the indirect effects of such interventions. However, the use of dynamic models in health technology assessment is relatively new, and these models require inputs from a multidisciplinary standpoint, to ensure that they offer an accurate method for analysis.

The use of these models is becoming increasingly recognised by the international community. The World Health Organisation, for instance, has developed a set of guidelines for economic evaluations...
of immunisation programs, with a chapter outlining a clear justification for the use of static or dynamic models for this purpose\(^6\). Moreover, the ISPOR-SMDM Modelling Good Research Practices Task Force recently published a set of good practices that should be adopted when using dynamic transmission models for economic evaluations\(^7,9\). This paper hopes to add to this growing body of evidence and analysis by providing a clear summary of the concept behind dynamic transmission models and by outlining suggestions for the appropriate use of the dynamic approach when evaluating health technologies and interventions for infectious diseases.

**Principle of dynamic transmission model**

Static models and dynamic models differ predominantly through the way that they use infection rates (force of infection\(^1\)). In static models, infection rates are either constant over time or vary according to personal characteristics such as age. In contrast, the infection rates used in dynamic models depend on the contact pattern, transmissibility, and distribution of the infected population over time, meaning that the infection rate is not constant.

Health technologies that are used in the treatment or prevention of infectious diseases—such as vaccinations—provide individuals with protection against specific infections. Vaccinations also reduce the rate of infection in the community within which those who are vaccinated live, by limiting the risk of an outbreak of the infectious disease. This is true only as long as a critical portion of a given community has been immunized against a contagious disease, which allows the spread of the infectious disease to be contained. This is known as “herd immunity”. An example of this can be seen in the eradication of smallpox, which resulted from a global vaccination program\(^10\). These kinds of indirect effects can be seen not only in vaccination programs but also in health interventions such as the screening and treatment for sexual transmitted diseases\(^9,10\). However, not all indirect effects resulting from health interventions benefit society. For instance, vaccination programs may also result in negative effects for a society. This can be seen in one case from the UK, where a pneumococcal conjugated vaccine program caused a pathogen strain replacement in the community within which it was implemented as well as in the case where a childhood vaccination program for the varicella-zoster virus resulted in age-shifting, where the disease began breaking out in other age group populations\(^11,12\).

Dynamic transmission models are useful for evaluating technologies that affect disease transmission within the interested population or ecology of the pathogen. If such interventions or technologies do not affect the force of infection, a static model may be used instead.

**Model structures**

In any economic evaluation, modeling the disease pathway accurately is a crucial step in determining which model is most appropriate. Depending on the natural history of the infection, various “compartments”—more widely known as health states—are defined for the model. These are usually set as susceptible, pre-infectious, infectious, and recovered/immune. The susceptible (S) state represents those who are at risk of infection. The pre-infectious (E) state occurs when an infected person is not yet able to pass the disease on to others but may become infectious in the future, while the infectious (I) state is the state in which a person becomes infectious. The recovered/immune state (R) is where a person becomes immune or has recovered from the infectious state.

As mentioned earlier, to determine which model is most appropriate, a full understanding of the disease pathway is essential. For example, in a case where an individual becomes infected with a disease and is re-infected again, a Susceptible-Infected-Susceptible (SIS) model may be deemed most appropriate. However, if the disease is curable, then a Susceptible-Infected-Recovered (SIR) model may be more appropriate. Examples of disease transmission in infectious diseases are shown in Fig. 1. Transmission models shown in Fig. 1 are appropriate when modeling a disease where transmission is between humans. However, when dealing with diseases that involve non-human hosts (parasites) such as Helminths and Anthropods, a model representing a disease pathway within the non-human hosts may also be necessary\(^13\).

Once a disease pathway has been developed, the next step in developing the model is to identify the pathway of the technology of interest. Where a vaccination program is introduced, those who were previously recognised as susceptible and had been vaccinated would shift to the immune stage, given the vaccine efficacy and coverage. Fig. 2 illustrates the pathway of population shifting from the susceptible compartment to the recovered/immune compartment in the SEIR model.

Once the structure of the transmission disease has been determined, the differential equations to
is the rate at which susceptible individuals become infected per unit time at time $t$, also known as the force of infection. Similar to the development of a typical static model, the identification of input parameters is important. Specific parameters used in the dynamic transmission model are shown below:

**Force of infection, $\lambda(t)$**

The force of infection is a function of the number of infected individuals in the population and its contact rate between the infected individuals and susceptible individuals:

$$\lambda(t) = \beta I(t)$$

Where $I(t)$ denotes the number of infected individuals at time $t$, and $\beta$ denotes the ability of successful infections (per capita rate), which depends on the spread of disease. The ability of successful infections depends on the route of transmission (i.e., whether respiratory, sexual, or vector-borne) and the density of the population. For example, a respiratory infection will spread at a higher rate in urban areas compared to rural areas (higher value of $\beta$), and a disease is transmitted by mosquitoes, the infection rate is likely to be higher in slum areas than in open-air areas. Other factors such as the age of the population, the distances and frequency of population travel within the area also play an important role. Therefore, it is difficult to obtain $\beta$ directly. Instead, it is calculated using the formula below, given that the individual in the population mixes randomly:

$$\beta = \frac{R_0}{ND}$$

$N$ is the total population and $D$ is the amount in time for disease spread; $R_0$ is the basic reproduction number, meaning the average number of secondary infectious persons resulting from one infectious person in the susceptible population. If the basic reproduction number is greater than 1, it describes the point where the epidemic starts. Fig. 3 illustrates the basic reproduction number equal to 5.

When introducing a national vaccination program, herd immunity can—theoretically at least—be reached, given the size of vaccine coverage. This coverage is set as a key target of the program and referred to as the herd immunity threshold (HIT). It is related to $R_0$ and can be calculated as follows:

$$\text{HIT} = 1 - \frac{1}{R_0} = \frac{R_0 - 1}{R_0}$$

Fig. 1 Common model structures for infectious diseases between human hosts.

Fig. 2 The disease transmission of an infectious disease when considering vaccination.

represent the transition of the population between each compartment are then defined in order to calculate the population in each compartment at a given point in time. Afterwards, the costs and outcomes of each event are calculated. An example of the set of differential equations for a SEIR model with effect from vaccination in birth cohorts is listed below:

$$\frac{dS(t)}{dt} = b(1 - \nu)N(t) - \lambda (t) S(t) - mS(t)$$

$$\frac{dE(t)}{dt} = \lambda (t) S(t) - fE(t) - mE(t)$$

$$\frac{dI(t)}{dt} = fE(t) - rI(t) - mI(t)$$

$$\frac{dR(t)}{dt} = b\nu N(t) + rI(t) - mR(t)$$

where,

$S(t)$ denotes the number of susceptible individuals at time $t$; $E(t)$ denotes the number of individuals in the latent period of infection at time $t$; $I(t)$ denotes the number of infectious individuals at time $t$; $R(t)$ denotes the number of individuals who recovered from the disease or were immune at time $t$; $N(t)$ denotes the total population of interest at time $t$; $b$ denotes the birth rate, or the rate at which individuals enter the specific population; $m$ denotes the death rate, or the rate at which individuals exit the specific population; $\nu$ denotes the proportion of individuals receiving vaccination; $f$ denotes the rate of onset of infectiousness; $r$ denotes the recovery rate; and $\lambda (t)$ is the rate at which susceptible individuals become infected per unit time at time $t$, also known as the force of infection.
Event rates

The event rate is the rate at which the population shifts between each compartment. In general, the rates are sometimes fixed. For example, the pre-infectious period is 2 days; therefore, the pre-infectious rate is equal to 0.5 per day. Therefore, setting the time step-size is very important as it will have an effect on the calculation, similar to setting the cycle length when using a Markov model. Infectious diseases usually have a small unit of time step-size, either in days or months unless they are for long-term infections. It is worth noting that the risk or the probability—the proportion of such event in each time step—should be used technically. However, the risk is approximately equal to the rate when the rate is small\(^{(14)}\). The relationship between rate and risk is shown as follows:

\[
\text{risk} = 1 - e^{\text{rate}}
\]

Important elements that need to be considered when conducting an economic analysis of interventions for infectious diseases using a dynamic transmission model

Before using the number of infectious populations resulting from a dynamic model, the model needs to be validated through the use of local data such as the size of the infectious population at a given point in time. However, within developing countries, many epidemiological data like these are fragmented. Moreover, nationally-represented epidemiological data are sometimes collected passively from compulsory notifications and reporting from provincial hospitals and may require further tests to confirm the types of strains to which they refer. In addition, for some infections where hospitalization is not usually required, the data reported in the national surveillance may be understated. Active surveillance is, therefore, preferable. However, active surveillance usually requires additional resources that are often unavailable, and the data themselves may not be widely applicable, given that any information gathered may only be relevant to areas within which the samples have been collected.

An example of the limitations of available data can be seen with dengue fever in Thailand. The surveillance system in Thailand does not require laboratory confirmation to report dengue infection cases, and recent research revealed that passive surveillance is likely to underestimate incidence of dengue fever significantly when compared to rates revealed by active surveillance. Fig. 4 illustrates another example of the comparison between observed and predicted cases of influenza in Thailand; this example was chosen because national data are available.

Result presentation

When presenting results of economic evaluations of programs concerned with preventing or treating infectious diseases, researchers should ensure that all transmission dynamics, including the incidence or prevalence of the disease, should be reported over time. If the particular intervention or technology is able to cause herd immunity within the population, the effects of the technology should be presented separately as direct and indirect effects (see example in Fig. 5). Where applicable, the results should also outline any relevant information specific to the infectious disease, such as strain replacement or the development of drug resistance. A sensitivity analysis should also be conducted for important input parameters. For transparency, the initial values and the set of differential equations should be provided. For more information on appropriate results presentation, please refer to the guidelines of the ISPOR-SMDM Modeling Good Research Practices Task Force\(^{(7,8)}\).

Although the demand for the use of dynamic transmission models in evaluations of health technologies for infectious disease is on the rise, they have not yet become part of the traditional toolbox of health economists, due in part to the fact that the results are complex and difficult to analyze, requiring extensive
whether a dynamic model is needed. For a dynamic model to be used in the economic evaluation, at least one of criteria below should be fulfilled:

1. The technology has an effect on the force of infection in the studied population.
2. The technology has an effect on the ecology of pathogen, such as strain replacement and antibacterial resistance.
3. The technology has an effect on the pathogenicity or the transmissibility of the disease, justified by the infection rate among other age groups.
4. The comparators of the interested technology have an impact on the infection rate.
5. There is a need for observing the incidence of the disease.

If none of the above criteria are met, a static model should be used. Moreover, a static model can also be used when existing studies with similar context show that certain technologies may represent good value-for-money. This is applicable for technologies known to have positive externalities, as the dynamic perspective will provide a more favourable incremental cost-effectiveness ratio.

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

2. Jit M, Brisson M. Modelling the epidemiology of infectious diseases for decision analysis: a primer.


แนวทางในการประเมินความคุ้มค่าต้านสุขภาพของโรคติดเชื้อ

ปัจจุบัน ที่ว่า

ในปัจจุบันมีการนำเอาวิธีการตัดสินแบบเชิงคัดศาสตร์หรือแบบจำลองคัดศาสตร์ที่มีออกมามากในวงการที่เป็นการประเมินความคุ้มค่าของการใช้ยา การตัดการและการรักษาโรคติดเชื้อ
ซึ่งเป็นการบันทึกผลโดยไม่มีผลกระทบการบันทึกผลเสียหรือเสียหาย ซึ่งมีวิธีการตัดสินแบบเชิงคัดศาสตร์ในปัจจุบันบางอย่าง
หรือสอดคล้องกับวิทยาของการรักษา บทความนี้เป็นการทบทวนความรู้ในรายละเอียดที่สำคัญที่ส่งเสริมรายละเอียดในรายละเอียดต่างๆ ของการประเมินความคุ้มค่าทางสุขภาพของโรคติดเชื้อ พร้อมทั้งให้ข้อมูลเฉพาะเจาะจงการประเมินความคุ้มค่าทางสุขภาพของโรคติดเชื้อ
Economic Evaluation of Screening for Disease

Naiyana Praditsitthikorn BSc(Pharm), PhD*,**

* Health Intervention and Technology Assessment Program, Ministry of Public Health, Nonthaburi, Thailand
** Bureau of AIDS TB and STIs, Department of Disease Control, Ministry of Public Health, Nonthaburi, Thailand

There is an increasing number of attempts to provide screening test or other interventions to prevent individuals from having disease. For the success of screening programs, appropriate screening tests should be provided to the right people, at the right time and at the desirable rate. Therefore, conducting a disease model and economic evaluation would help decision-makers and stakeholders to assess the surrounding factors and ensure the successfulness of the program prior to an implementation in the real setting. Because the evaluation of screening tests is particularly specific, this chapter aims to conclude that the needed information and examples must be made clearer.

Keywords: Economic evaluation, Screening

J Med Assoc Thai 2014; 97 (Suppl. 5): S94-S101
Full text. e-Journal: http://www.jmatonline.com

The Commission on Chronic Illness Conference on Preventive Aspects of Chronic Disease, held in 1951, defined screening as “The presumptive identification of unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly. Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic. Persons with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment”(1). In general, there are three types of screening programs: 1) Mass or population-based screening indicated for the large-scale screening where no selection of population groups is made (e.g. thyroid stimulation hormone screening in newborns, human immunodeficiency virus testing in pregnant women); 2) Surveillance screening, a term used in the sense of a long-term process where screening examinations are repeated at time intervals (e.g. routine screening for health workers, screening of lead and heavy metals in people living in industrial areas); 3) Opportunistic screening, which refers to tests offered to people who are being examined for other reasons (e.g. providing blood pressure and body mass screening every time patients visit their physician for general consultation or unrelated health problems).

Everyone knows the advantage of screening in which a member of a defined population may not necessarily perceive their risk or are already affected by disease or its complications, and are offered information or further tests and appropriate treatment to reduce their risk and/or possible complications arising from the disease or condition. However, there are many limitations and concerns on achieving effective and appropriate screening programs. Wilson and Jungner, 1968 developed 10 criterion that must be fulfilled to guide the case-finding programs(2).

1) The condition sought should be an important health problem.
2) There should be an accepted treatment for patients with recognized disease.
3) Facilities for diagnosis and treatment should be available.
4) There should be at a recognizable latent or early symptomatic stage.
5) There should be a suitable test or examination.
6) The test should be acceptable to the population.
7) The natural history of the condition, including development from latent to declared disease, should be adequately understood.
8) There should be an agreed policy on whom to treat as patients.
9) The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
10) Case-finding should be a continuing
process and not a “once and for all” project.

In conclusion, before a screening program is implemented, it should have appropriate knowledge of disease characteristics and treatment, knowledge of screening test and cost considerations. This chapter will provide the information needed if conducting an economic evaluation study that can be illustrated in four topics.

1. Knowledge of disease

The natural history of disease should be appropriately known either before or after the disease occurred. This can be divided into pre-clinical and clinical phases (Fig. 1). The pre-clinical phase is the period from the biological onset of disease to the onset of clinical manifestations of the disease (A to S). During the pre-clinical phase, the condition is asymptomatic but has an interval that is detectable via a screening test (B to S). Hence, the benefit of screening is that the detection at the pre-clinical phase of disease results in a better prognosis than therapy given after symptoms develop; otherwise, the screening is not necessary.

In case of cervical cancer screening programs, it is based on the premise that invasive cervical cancer results from the progression of pre-invasive precursor lesions called cervical intraepithelial neoplasia (CIN), which progress from mild (CIN-1) to moderate (CIN-2) to severe (CIN-3) and then to cancer. It appears that CIN progresses to cancer over a prolonged period usually 7 to 20 years and is asymptomatic. This long natural history provides the opportunity for screening during the pre-invasive phase. Because mild lesions (CIN-1) would spontaneously recover and never progress to invasive cancer, these women will be monitored rather than treated. Therefore, only the women at greatest risk (CIN-2/3) for developing invasive cancer need to be treated.

Moreover, the availability of disease epidemiological statistics should be checked, for instance, incidence and prevalence of the precursor of screening among targeted populations, disease morbidity and mortality rate. For a complicated condition or disease that involves through several health states over time, it requires the data of probability of changing health conditions over the time period (called state transitional probability). This can be identified from many sources of information, e.g. observational study or clinical trial. In Thailand, there are some statistical databases that can be publicly accessed; e.g. the Population and Housing Census by the National Statistical Office, data of disease surveillance by the Bureau of Epidemiology, Cancer Registry by the National Cancer Institute, Burden of Disease by the International Health Policy Program and Health Statistics by the Bureau of Policy and Strategy.

2. Knowledge of screening test and its effectiveness

The effective screening program does not comprise only the accuracy or validity of the test but also the managerial factors of the program. For a successful program, the appropriate screening test should be provided at the right time to the right people and at the desirable rate. The benefit of developing a decision model is to help identify whether a screening test, targeted population and screening interval should be prior to the real program’s being implemented, also how the cost and consequences will occur when the situation changes.

Effectiveness of screening tests

There are multi-factors that will affect the effectiveness of screening programs which need to be considered, e.g. target population, frequency of screening, acceptance and loss to follow-up rate.

To be appropriate for screening, the pre-clinical phase of disease should have a high prevalence among the target population for screening. In some cases, providing a screening test to the whole population would be too costly. Consequently, targeting high-risk populations can increase the prevalence of detectable, pre-clinical phase and number of cases detected. High-risk population may refer to a group of people who has specific demographics, e.g. age, gender, race, family history, or having risk behavior. In spite of recurrence of pre-clinical phase, some conditions or disease should be rescreened regularly. Being aware that short intervals of screening may incur high costs compared to a lesser number of case detections, while a long interval may result in being too late to prevent people from falling into the clinical phase of disease. By the way, some conditions permanently occur and only need screening once in a lifetime.

As mentioned above, the effectiveness of screening relies more on multiple factors then would be considered, relatively: for example in cervical cancer screening programs in low-income countries because of very high incidence of disease and limited number of resources, providing a screening test only once in a women’s lifetime at the high level of coverage is the most cost-effective option compared to several
screenings, but at very low coverage rate. Fig. 2 illustrated how to apply an acceptable rate of screening and lost to follow-up rate into the decision tree model.

**Accuracy of screening test**

Ideally, a screening test should be accurate and neither miss cases nor falsely identify healthy people as having disease. The assessment of a screening test’s accuracy depends on the comparison of screening results to a reference standard, which commonly refers to sensitivity, specificity\(^9\,10\).

Sensitivity or true positive rate is the proportion of truly diseased people in the screened population who are identified as diseased by the screening test. Sensitivity is a measure of the probability of correctly diagnosing a case or the probability that any given case will be identified by the test written in conditional probability form: \(P(T^+|D^+).\)

In general, a highly sensitive test should be selected when the consequences of missing a disease would be a bad outcome.

Specificity or true negative rate is the proportion of people without the disease who are identified as such by the screening test. Specificity is a measure of the probability of correctly identifying a non-diseased person with a screening test written as \(P(T^−|D^−).\) A highly specific test should be selected when false positive results can substantially harm the patient.

Sensitivity and specificity are dependent on cut-off value to which the test is positive. Nevertheless, they are not useful for clinicians because generally it is not known whether the patient has the disease. Unlike sensitivity and specificity, the positive predictive value (PPV) and negative predictive value (NPV) are dependent on the prevalence of disease and useful for clinicians to determine probability of disease among positive test results and probability of no disease among negative test results. They are defined as:

Positive predictive value is the probability that a people with a positive test results on the screening test truly has the disease written as \(P(D^+|T^+).\)

Conversely, Negative predictive value, is the probability that a people with a negative results on the test truly does not have the disease written as \(P(D^−|T^−).\)

The relationships are often shown in a fourfold table in which the letter \(a, b, c\) and \(d\) represent the quantities. In addition, basic formulas generally used in an assessment are given in Table 1.

Because very few tests are both highly sensitive and highly specific, two or more tests may be performed either in parallel (sometimes called simultaneous testing) or in series (two-stage testing). The former means tests are applied at the same time and interpreted together. The latter describes the results of the first test to determine whether the second test is performed at all; this method has the advantage of avoiding unnecessary tests, but the disadvantage of delaying diagnostics and treatment by lengthening the diagnostic testing period.

Table 2 illustrated the decision rules for two combined test. Performing in parallel and in a series of tests can be interpreted in two ways: Firstly, what is called “the OR rule”, which yields a positive result if either test is positive or a negative result if all tests are negative. Consequently, the sensitivity of the combined result; denoted by \(SE_c\) is higher than that of either test alone, but the combined specificity denoted by \(SP_c\) is lower than that of either test.
Table 1. Classification of screening outcomes

<table>
<thead>
<tr>
<th>Screening test results</th>
<th>True diagnosis</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Diseased</td>
<td>Not diseased</td>
</tr>
<tr>
<td>Positive</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Negative</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>Total</td>
<td>a + c</td>
<td>b + d</td>
</tr>
</tbody>
</table>

a = Diseased individuals detected by the test (true positives)
b = No diseased individuals positive by the test (false positives)
c = Diseased individuals not detectable by the test (false negatives)
d = No diseased individuals negative by the test (true negatives)

Sensitivity (SE) or $P(T^+|D^+)$ is $\frac{a}{a+c}$
Specificity (SP) or $P(T^-|D^-)$ is $\frac{d}{b+d}$
Prevalence (p) is probability of having disease; $P(D^+) = \frac{(a+c)}{(a+b+c+d)}$

Positive predictive value (PPV) or $P(D^+|T^+)$ is $\frac{a}{a+b}$

written in probability notation:
$$P(T^+|D^+)P(D^+) \text{ or } SE \times p$$
$$P(T^+|D^+)P(D^+) + P(T^+|D^-)P(D^-) \text{ or } [(SE \times p) + [(1-SP) \times (1-p)]]$$

Negative predictive value (NPV) or $P(D^-|T^-)$ is $\frac{d}{c+d}$

written in probability notation:
$$P(T^-|D^-)P(D^-) \text{ or } SP \times (1-p)$$
$$P(T^-|D^-)P(D^-) + P(T^-|D^+)P(D^+) \text{ or } [(SP \times (1-p)) + [(1-SE) \times p]]$$

Probability of a positive test result or $P(T^+)$ is $\frac{(a+b)}{(a+b+c+d)}$
written in probability notation:
$$= P(T^+|D^+)P(D^+) + P(T^+|D^-)P(D^-) \text{ or } (SE \times p) + [(1-SP) \times (1-p)]$$

Probability of a negative test result or $P(T^-)$ is $\frac{(c+d)}{(a+b+c+d)}$
written in probability notation:
$$= P(T^-|D^+)P(D^+) + P(T^-|D^-)P(D^-) \text{ or } [(1-SE) \times p] + [SP \times (1-p)]$$

The second rule, called “the AND rule”, yields a positive result only if all tests are positive and a negative result if either test is negative. Therefore, the specificity of combined results is higher than either test alone, but the combined sensitivity is lower than that of either test.

Performing tests in series is remarkably cost-efficient when screening for rare conditions and often used when the second test is expensive and/or risky. With the OR rule, if the first test is positive, the diagnosis is positive; otherwise, the second test is performed. If the second test is positive after a negative first test, then the diagnosis is positive; otherwise, the diagnosis is negative. The OR rule, leads to a higher overall sensitivity than either test in itself. Under the AND rule, if the first test is positive, the second test is performed. If the second test is positive, the diagnosis is positive; otherwise, the diagnosis is negative. The AND rule then leads to a higher overall specificity than either test by itself(11).

In spite of developing a decision model to decide on the appropriate screening program, it is important to define the research question, competing strategy, data availability then appropriate methodology and model selection; e.g. decision tree, Markov model, discrete event simulation and dynamic model. This chapter will demonstrate only the decision tree model, which is frequently used to undertake economic evaluation of a screening program. In constructing the decision tree, events identification and ordering should be considered. In general, the order of the events usually follows the sequence of events over time according the logical progression of the decision pathway. In addition, the way to ordering events in decision tree model of any screening or diagnosis tests can be structured by either process...
Table 2. Decision rules and formula of combined test accuracy, assuming two screening tests performed

<table>
<thead>
<tr>
<th>Decision rules</th>
<th>Accuracy of combined test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test A</td>
<td>Test B</td>
</tr>
<tr>
<td><strong>Parallel or simultaneous testing, OR rule</strong></td>
<td></td>
</tr>
<tr>
<td>Either test positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td><strong>Parallel or simultaneous testing, AND rule</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Either test negative</td>
<td>Negative</td>
</tr>
<tr>
<td><strong>Series or sequential testing, OR rule</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>Not performed</td>
</tr>
<tr>
<td>Negative</td>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
<td>Negative</td>
</tr>
<tr>
<td><strong>Series or sequential testing, AND rule</strong></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>Positive</td>
</tr>
<tr>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Negative</td>
<td>Not performed</td>
</tr>
</tbody>
</table>

SE<sub>a</sub> = sensitivity of test A alone; SE<sub>b</sub> = sensitivity of test B alone; SE<sub>c</sub> = overall sensitivity; SP<sub>a</sub> = specificity of test A alone; SP<sub>b</sub> = specificity of test B alone; SP<sub>c</sub> = overall specificity

Fig. 3 Structure of the decision tree: process ordered.

Fig. 4 Structure of the decision tree: disease status.

The simple decision trees illustrated in Fig. 3 and 4 showing the way to apply screening test’s accuracy and relationships obtained for four screening outcomes: i.e. true positive, false negative, true negative and false positive. The first example of ordering is by process ordered (assuming full coverage of screening); if not, see section Effectiveness of screening test and ordered (Fig. 3) or by disease status (Fig. 4).

p(T+) = (sensitivity x prevalence) + [(1-specificity) x (1-prevalence)]; p(T−) = 1-p(T+); p(D+) = prevalence of the disease; p(D−) = 1-prevalence; p(D+[T+] = PPV; p(D−[T+] = 1-PPV; p(D+[T−] = 1-NPV; p(D−[T−] = NPV

P(D+) = prevalence of disease; P(D−) = 1-prevalence; P(T+[D+]) = sensitivity; P(T−[D+]) = 1-sensitivity; P(T+[D−]) = 1-specificity; P(T−[D−]) = specificity.

The screening initially divided people into two groups according to the test results. Test positive group will be referred to further investigation and treatment then subsequently divided into true positive and false positive groups by which clinical decisions are taken. Under the test negative group, there will be no further investigation, the people will be divided into true negative and false negative groups according to clinical status. Each subsequent branch required probability of events occurred; e.g. probability of positive test
results ($P(T+)$), probability of disease conditional on positive test results ($P(D+|T+)$ or PPV).

A substitute model is to start with the assumption that there is a proportion of people with disease that show a prevalence of disease. Under diseased branch, screening then has subsequent probabilities of test positive ($P(T+|D+)$ or sensitivity) and test negative ($P(T−|D+)$ or 1-sensitivity). This approach is much simpler for applying the probability of events especially when the underlying disease and event pathways are complicated, but less intuitive for clinicians.

Either conducting the decision tree by process ordered or according to disease status, the four screening outcomes will be the same. The true positive group is obtained the most benefit from early detection according to screening program while the false positive group is more costly and may be harmful for having undertaken unnecessary investigation and treatment. The disadvantage group of screening is from a false negative result, because the test indicates no disease hence the people, who actually have disease, are delayed in receiving treatment. The true negative group is affected only as to the cost associated with the screening program$^{[9,12]}$.

3. Validation of a decision model

Once the decision model has been constructed and analyzed, it is important to assess its validity and check for its consistency. In brief, face or descriptive validity needs to check whether model structure, assumptions and results are reliable and can be explained accordingly. Internal validity relates to the logic of the model and consistency of the model inputs to its outputs. Predictive validity refers to the ability of the model to make accurate predictions of future events$^{[9]}$. One of the effective ways to illustrated internal and predictive validity is to compare clinical outputs predicted from the model against external data from real population or clinical trial. The valid model should draw results under particular situation and fit well with the real data.

Fig. 5 showed the example from an economic evaluation of cervical cancer screening (Praditsithikorn N et al. 2011). Because human papillomavirus (HPV) infection has a causal relationship with the occurrence of cervical cancer, the chances of HPV infection rises in the teenage population (Fig. 5A), and infected women may develop cervical cancer in older age (Fig. 5B). Moreover, in Thailand before 2005, there was no population-based cervical cancer screening only opportunistic screening that identified a 10% coverage rate. As a result, the prediction of cervical cancer incidence derived from the model under specific situation of 10% coverage of screening was then compared with population data in 1999. The results show that the model tends to be valid enough for predicting disease outcomes.

4. Concerning on results presentation

Informative results will be helpful to decision-makers and stakeholders to decide the most cost-effective program. Because economic evaluation of screening for disease is involved, some kind of comparison between alternative screening test, administration programs or treatment pathways, is needed. Therefore, the two (or more) options to compare, and two dimension of outcome, i.e. cost and health outcome, should be presented appropriately. Some concerns are discussed below:

- Cost outcome should be reported in a disaggregated way. Regarding the screening program,
it aims to prevent people, through early detection of disease, the cost and its consequence in terms of the cost of prevention apart from the cost of disease treatment should be provided.

- Health outcome of each alternative can be calculated by using many different types of measurement units. For instance, two screening options could be compared in terms of number of case preventions, the number of deaths averted and the number needed to screen for a one-case prevention. This approach of surrogate outcome analysis is useful and intuitive for clinicians.

- Final health outcome or quality-adjusted life year (QALY) is widely used and comparable across different areas of health problems. Consequently, it is generally used by decision-makers to help compare and allocate the resources.

- In cost-effectiveness and cost-utility analysis, the calculation of effects between two options are in terms of difference in cost and difference in health outcome should be presented in the form of ratio, namely incremental cost-effectiveness ratio (ICER).

- In case of more than two options, it is not necessary to calculate all possible ICERs for every pair of options, but only the efficient options.

- For decision-making, the variation of results based on major administrative parameter changes as well as budget constraints should be presented to help facilitate program planning.

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
การประเมินความคุ้มค่าทางสาธารณสุขของมาตรการคัดกรองโรค

นัยนา ประดิษฐ์สิทธิ์

เปรียบเทียบหลายหน่วยงานที่มีการทำความเสียหายต่อการส่งเสริมสุขภาพและป้องกันโรคมีอยู่ซึ่งรวมทั้งหน่วยงานที่มีการคัดกรองโรค อย่างหนึ่งที่มีความสำคัญของการคัดกรองซึ่งถูกนำมาเปรียบเทียบคือ ความรุนแรงของการคัดกรอง เทคนิคการคัดกรอง การบริการจัดการโรค เป็นต้น ดังนั้นเพื่อให้เรามีความรู้เกี่ยวกับแผนการคัดกรองอย่างมีประสิทธิภาพและได้ประโยชน์สูงสุด การประเมินความคุ้มค่าทางสาธารณสุขของการคัดกรองแบบต่างๆ จำเป็นต้องพิจารณาการเลือกแบบความเหมาะสมในการปฏิบัติและเปรียบเทียบผลการที่จะได้รับทั้งทางสุขภาพและเศรษฐกิจการ เพื่อรองรับการคัดกรองโดยที่จะนำไปปฏิบัติจริง เนื่องในหนึ่งมีวิธีคัดกรองโรคที่จะมีผลถึงความคุ้มค่าของมาตรการคัดกรองใดๆ
The Economic Evaluation of Medical Devices: Challenges

Pritaporn Kingkaew BPharm, MSc*, Yot Teerawattananon MD, PhD*

* Health Intervention and Technology Assessment Program (HITAP), Nonthaburi, Thailand

While many of the principles that guide the economic evaluation of medical devices are somewhat similar to those that guide the evaluation of other health technologies, most outline a methodology that focuses on pharmaceutical products rather than providing specific guidance for medical devices. Given the wide range of technologies and can be used for many purposes, conducting an economic analysis for medical devices is not straightforward. The cost and effectiveness of a given technology may depend on many factors. The objective of this paper is to provide a summary of issues that need to be addressed before undertaking an economic evaluation of medical devices and to outline a number of suggested approaches for undertaking an economic evaluation of medical devices.

Keywords: Technology assessment, Cost-benefit analysis, Cost, Effectiveness, Medical device

Medical devices are many and diverse. The World Health Organisation defines a medical device as “an article, instrument, apparatus, or machine that is used in the prevention, diagnosis, or treatment of illness or disease, or for detecting, measuring, restoring, correcting or modifying the structure or function of the body for some health purpose. Typically, the purpose of a medical device is not achieved by pharmacological, immunological or metabolic means(1). There are several nomenclature systems for medical devices in use around the world. The Global Medical Devices Nomenclature System (GMDN) and the Universal Medical Devices Nomenclature System (UMDNS) are the two most common(2). In Thailand, there is no standard nomenclature system or definition; however, the Medical Device Acts BC 2551 does give a definition of medical devices similar to that of the WHO(3).

In recent years, economic evaluations of healthcare interventions have become highly sought after by Thai decision-makers(4). An economic analysis is one of the policy tools that can be used to justify the inclusion of a given healthcare intervention in healthcare benefit packages. According to Section 22 of the Thai Medical Device Acts BC 2551, upon the production or import of pre-defined medical devices, all manufacturers should submit a request to the licensing authority to assess the medical device in terms of effectiveness, quality, standard, safety, and efficiency. They should also assess the socioeconomic impact, value-for-money, suitability, and inclusiveness of the device introduction and examine equitable uses of the devices(3). This legislation indicates growing concern among Thai policy-makers regarding the cost-effectiveness of medical devices.

While the evaluation of medical devices is similar to that of other health technologies, there are important distinctions. Despite this, the available guidelines are either very general or focus only on pharmaceutical products; no guidance has yet been developed that specifically addresses the evaluation of medical devices(5). This may be due to the fact that the regulatory framework for medical devices is less rigorous than the regulatory framework for pharmaceutical products(6). Moreover, the category of medical devices is broad, encompassing as it does prevention, diagnosis, treatment, storage, detection, measurement, and inspection. In addition, the function and duration of usage for any given medical device can also vary significantly, from single disposable devices such as pregnancy tests, HIV tests, rubber gloves, or condoms, to devices that have applications for a longer duration such as large equipment that is used for diagnostic purposes or implant devices.
This diversity can make it difficult to formulate guidance that applies to all medical devices, a challenge that is compounded by the fact that the effectiveness of a medical device can vary significantly depending on a number of factors, including device-operator-patient interaction and device adjustment or maintenance; conducting an analysis of medical devices is not straightforward. Given these specific complexities, there are certain important aspects that should be taken into consideration when conducting a cost-effectiveness evaluation of a medical device. The objective of this paper is to provide an overview of these issues and to outline the best approach when conducting an economic analysis of medical devices, especially in limited-resource settings.

**Consideration for the efficacy of medical devices**

**Quality of the study**

In the hierarchy of research designs, the randomised control trial (RCT) is regarded as one of the most reliable sources of data. However, for medical devices, RCTs are often not feasible for a number of reasons. For instance, concealment or blinding is often not practical for medical devices, which leads to biased results. In addition, randomization is often not possible because many devices require an invasive procedure, which requires consent from subjects, and because medical devices are used as standard practice, which makes randomization unfeasible because of ethical issues. Moreover, even when RCT is applicable to medical device evaluation, the analysis results may still be limited in terms of target population, sample size, and the time scale of monitoring and evaluation of the technologies. For these reasons, effectiveness evaluations of medical devices are rarely based on experimental studies with randomization and control; instead, observational studies are usually relied upon.

**Multiple applications**

Medical devices that are used as screening or diagnostic tools commonly have multiple applications. For example, Positron Emission Tomography/Computed Tomography, or PET/CT, is a tool used for diagnosis and follow-up of cancers such as cervical cancer, colon cancer, oesophageal cancer, non-small cell lung cancer, lymphoma, melanoma, and ovarian and thyroid cancer. Therefore, in order to fully assess the benefit of PET/CT, all of its various applications and indications need to be taken into account. Although other health technologies, including pharmaceutical products, are also subject to these difficulties, it is usually much easier to differentiate the effectiveness by indication for pharmaceuticals than it is for devices. This is, in part, because it is challenging to differentiate and quantify the benefit and cost per application for medical devices that have multiple applications, which means that the benefit of a given device may well be underestimated if all of its various benefits are not taken into account. As such, all benefits should be taken into account when conducting a cost-effectiveness analysis of medical devices with multiple applications.

**Intermediate and final outcomes**

Since a cost-utility analysis is one of the standard tools typically recommended when making coverage decisions for benefit packages in tax-based national health insurances, it is important that data on the outcomes of the technology under consideration, given in terms of quality-adjusted life year (QALY) or disability-adjusted life year (DALY) be available. However, for medical devices, this can be somewhat complicated, given that, in most cases, the efficacy or effectiveness of a medical device is reported in the form of intermediate outcomes, such as success rate, complication rate, procedure duration, diagnostic performance, and level of measurement. In this case, a cost-utility analysis can only be performed where a strong causal link between the intermediate and final outcome can be demonstrated, which is not very common. For instance, with medical devices that are used for screening or diagnosis, the outcome of the device will vary depending on the episode of care of the patient, which will be determined by the diagnosis following use of the device. Consequently, to evaluate the real costs and benefits of medical devices, the entire pathway of care should be taken into account for all true positive cases, false positive cases (which can result in unnecessary worry for patients and their families), and false negative cases (which can result in a delay treatment or care).

**Device-operator interaction**

For medical devices that require personnel for operation or interpretation, the performance of the device to prevent, diagnose or treat a condition depends not only on the device itself, but also to a significant degree on the operator’s skill and experience-known as “device-operator interaction”. The higher the skill or experience of the operator, the greater the benefits of the device will be; this is known as a “learning curve”. This was demonstrated, for instance, in one multicentre study involving the USA...
and several European countries, which proved that the benefits of laparoscopic radical prostatectomy improved significantly in line with operator experience and skill\(^{(13)}\). The present study found that the positive surgical margin, which represents the efficacy of the surgical procedure, improved with the number of surgeries the operator had conducted, reaching a plateau at approximately 200 to 250 surgeries. Researchers conducting analyses of medical devices that rely on device-operator interaction should take into consideration the stage of development of the technology. For instance, a technology that is assessed in the early stages of development may well generate less positive results that it would if assessed at a later stage, as the operators are likely to have become more experienced.

**Incremental innovation**

While research and development of pharmaceutical products usually takes a decade or more, this is not the case for medical devices. Many devices have a short product life-cycle, approximately 1-3 years\(^{(13)}\), usually undergoing only minor incremental innovations such as those that result in a longer battery life or improved interactive functions that render the device easier to use. A good example of medical device evolution can be seen in the self-monitoring blood glucose meter which, after many iterations, now requires a smaller blood sample size, has reduced pain and test times, improved error detection routines and portability, and contains software integration to monitor blood glucose levels for physician facilitation. Hearing aids have also undergone frequent modification so that now there are many designs that can be adapted according to the user’s preference. These modifications have included the development of various different styles (behind-the-ear, in-the-ear, in-the-canal, completely in-the-canal, and body-worn hearing aids), different sound-amplifying systems (analog or digital hearing aids), and certain special features such as a directional microphone and telephone switching. As these medical devices demonstrate, medical devices tend to develop incrementally, with small but frequent stages of innovation. This means there is less incentive for the industry to invest in effectiveness studies of newer iterations unless there are significant changes that result from the modification.

**Factors affecting the efficacy of medical devices (“real world” factors)**

There are numerous factors that may contribute to the effectiveness of a given technology, including the training that is available for proper use and care or evaluation and monitoring. For individuals who have undergone a cochlear implantation, for instance, auditory/speech rehabilitation is required for users to be able to take full advantage of all the benefits of the technology; this means that it is not only the success rate of transplantation that affects how effective the cochlear implant can be said to be\(^{(14)}\). This can also be seen with hearing aids, where the effectiveness depends not only on the aid itself but also whether the patient regularly attends hearing aid fittings with their audiologist, and with self-monitoring blood glucose meters, where the effectiveness depends in part on the patient’s knowledge and ability to properly calibrate the device (an improperly adjusted device may generate incorrect glucose readings, resulting in hypoglycemic events).

Moreover, the effectiveness of the medical device will also depend on the context of analysis. While primary data are usually preferable, it is often unavailable or partial. For example, when the cost-effectiveness analysis of cochlear implantation was conducted, there were no data available regarding the success rate of the operation, so retrospective data collection of the rate was conducted instead\(^{(14)}\). In cases where an along-side clinical trial is implemented, the factors associated with the outcome may be controlled\(^{(15)}\). However, when evaluating the effectiveness of medical devices, researchers should try, as far as possible, to collect data from a context similar to that within which the real decision will be made. Where this is not possible, researchers should include real world factors (such as acceptance rate) when constructing models for the economic analysis of medical devices.

**Consideration for costing medical devices**

**Cost information and economy of scale**

Calculating the cost of a medical device can be relatively straightforward, for instance in the case of disposable devices. In contrast, evaluating the cost of medical devices that can be used numerous times or for various indications can be more challenging. In the context of an analysis where a device with a high level of technology is to be introduced into the health system, calculating the costs of start-up will be necessary. Two noteworthy examples of this are the magnetic resonance imaging (MRI) or positron emission tomography (PET) systems, which both require significant start-up costs, whether for screening or...
diagnostic options. This is because the proportion of the systems’ fixed costs is much higher than their variable costs. In this case, calculating fixed costs—where the cost of input for each unit of output stays the same in the short run—and variable costs—where the cost of input increases with each unit of output produced—are useful for determining the break even point. One recent study found that current PET and CT services were under utilised. On average, they found that PET and CT machines were used less frequently than they would need to be to break even (eight times per location per week), meaning that, at the current usage, providers needed to pay higher fix costs, resulting in a deficit. They calculated the fixed and variable costs of providing PET and CT services to be 14,773,376 baht per year and 20,714 baht per year, respectively, and suggested that this under utilization reflected patients’ unwillingness to use these services due to their expense. Clearly then, the calculation of fixed and variable costs of very expensive medical devices is an important part of ensuring that policies are developed in an integrated and informed way that ensures maximum benefit.

Continual change in price of the medical devices

As discussed, the fact that medical devices undergo continuous research and development, they have a much shorter product life-cycle compared to pharmaceutical products. This is especially true for electronic medical devices, where the product life-cycle is now around one to three years. One drawback of the continuous development of medical devices is that older technologies can sometimes appear more cost-effective since prices of older technologies drop considerably once a new technology has been introduced. For example, while drug-eluting stents were initially found to be cost-effective when they were first introduced into the United Kingdom, after four years this was no longer the case, because the price of the nearest treatment competitor, bare-metal stents, had decreased significantly in the wake of the introduction of the newest technology. This means that there are significant methodological difficulties with conducting economic evaluations for medical devices.

Guidelines for Health Technology Assessment in Thailand (second edition): Recommendations for the economic evaluation of medical devices

1. The costs and outcomes resulting from treatments of the screened or diagnosed diseases should be taken into account when evaluating the cost-effectiveness of screening or diagnostic medical devices. Even where various treatment options may be available, use of standard practice for management of the disease should be used as the comparator wherever possible, rather than no treatment. Moreover, the overall costs and outcomes resulting from appropriate practice should be used to model for the effectiveness of the device, rather than constructing a sub-model for each disease or condition.

2. When evaluating the cost-effectiveness of medical devices that require expert personnel for their operation, all data sources should be described and a justification for the use of effectiveness information in reference to that device should be clearly stated. Where secondary data are applied, a discussion should be included that explains whether information is appropriate for other analytical contexts.

3. Where the cost-effectiveness of the medical device under consideration depends on other factors such as distribution or usage, a feasibility analysis for such medical devices in the context of analysis is also recommended.

4. Where the value of the medical device under consideration depends on economies-of-scale, both fixed and variable costs should be obtained in order to calculate the break even point to determine the proper output value.

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


บทความของการประเมินความสุขภาพที่มีเกี่ยวกับการนั่งเล่นของเด็ก

ปัญญาระหว่าง อ. ศ. ศรีวัฒนา

การประเมินความสุขภาพและการกระทำสุจริตในการเล่นของเด็กไม่ว่าจะเป็นแพทย์โฆษณาหรือไม่คดีตาม
โดยทั่วไปได้เห็นกล่าวขานส่วนการประเมินที่เน้นกัน แนวทางการประเมินความสุขภาพจิตของเด็กไม่ว่าจะเป็นมากกว่าเครื่องมือแพทย์
อย่างไรก็ดีเครื่องมือแพทย์จะข้างต่างก็มีความที่ควรจะให้ความพิจารณา และหากที่จะการประเมินความสุขภาพของเด็กมีผลต่อแพทย์จึงมีผลต่อระบบสังคม
โดยการพิจารณาจากที่ดูผู้รักษารูปแบบ เครื่องมือแพทย์ที่มีหลายอย่างซึ่ง ปฏิสัมพันธ์ระหว่างเครื่องมือแพทย์และผู้เข้าร่วม การพัฒนาของเครื่องมือแพทย์
อย่างต่อเนื่อง ขึ้นอยู่กับผลของการพิจารณาผู้รักษารูปแบบของเครื่องมือแพทย์ที่ทำให้การประเมินความสุขภาพของเด็กมีผลต่อระบบสังคม
และระบบการยึดเครื่องมือแพทย์ที่มีการเปลี่ยนแปลงอย่างต่อเนื่อง ซึ่งเกิดจากการเปลี่ยนแปลงที่อยู่ต่อเวลาเป็นนั้น ทั้งนี้การความถูกต้องของนั้น
สำหรับการประเมินความสุขภาพสูงที่มีการจัดเก็บเครื่องมือแพทย์เพื่อเป็นประโยชน์แก่เด็กน้อยในอนาคต
Constructing a State-Transition Model for an Economic Evaluation of Cancer Treatments

Chulaporn Limwattananon MPharm, MSc, PhD*,
Supon Limwattananon MPHM, PhD*

*Faculty of Pharmaceutical Sciences, Khon Kaen University, Khon Kaen, Thailand

The paper gives an overview of the four fundamental elements that should be considered when constructing a Markov model of cancers, including outcome measures, health state transition, transitional probabilities, and model calibration. The construction of any model of this kind should begin by establishing transition to the death state. The probability of this transition can be estimated using overall survival data from clinical studies. Possible health states over a cycle are defined according to the natural history of diseases and treatment pathways. Validity of the constructed model is tested against real patient data and the parameters are adjusted until the survival results are consistent.

Keywords: Health state, Markov model, Solid tumor, Survival, Transitional probability

Among healthcare experts, there is growing recognition of the importance of "whole disease modeling" to help inform decisions regarding prevention and treatment(1). This paper provides an overview of the fundamentals of constructing a state-transition model (best known as a Markov model), to serve as a whole disease model for cancers and treatment pathways. Four related methodological issues are examined: 1) choice of outcome measures from clinical studies, 2) development of health state transition, 3) estimation of transitional probabilities, and 4) calibration of the constructed model.

Outcome measure

When building an economic evaluation model of cancer treatments, a crucial first step is deciding which outcome measures, derived from existing clinical and epidemiological studies, will be used in the model. In most randomised controlled clinical trials (RCTs) and comparative effectiveness research(2), overall survival (OS) or disease-free survival (DFS)-sometimes called progression-free survival (PFS)-are set as the primary end point, and treatment response is set as the secondary end point. For cancers where treatment is unlikely to increase patient survival, studies may choose to report the response rate instead.

For OS and DFS/PFS, the survival rate (measured as a 0–1.0 proportion or 0–100%) at various time points can be estimated using the Kaplan-Meier method. For ease of understanding, it is normally depicted as a survival curve, with the survival rate equal to 1.0 (or 100%) at time zero, which then declines over time as in the life table(3). Because DFS/PFS is a combined end point, consisting of surviving patients who have no progressive disease (PD), the DFS or PFS is always lower than the OS at any given time point. Median survival data (the duration of time that passes before half of the patients have died) and time to progression data should not be used directly for estimating a transitional probability of the model parameter.

Treatment response rate is usually derived from data from patients who at baseline have at least one cancerous cell with the longest diameter (LD) no shorter than 20 mm (or 10 mm based on spiral CT scans), and who afterwards are evaluated (i.e. excluding data from patients who drop out before evaluation). It is important to note that the response rate becomes an unreliable or unstable outcome measure in studies where the number of evaluated patients is far below the number at baseline.

For patients with solid tumors, an objective response (OR) to the treatment can be classified as either complete response (CR), where all cancer cells are found to have disappeared following treatment, or partial response (PR), where fewer cancer cells are found than before treatment or where the LD of the tumor is found to be reduced by at least 30%(4). Non-response (NR) outcomes can include stable disease (SD), where...
target lesions are found to have shrunk or enlarged by a certain degree (less than PR or PD) or progressive disease (PD), where the LD is found to have increased by at least 20% or where a new lesion is observed. PD indicates the second worst possible treatment outcome following death as a result of the disease. It is regarded as treatment failure, and often leads to death, which is referred to technically as the ‘absorbing health state’.

**Health state-transition**

A Markov model is a state-transition model[5] that simulates the transition of health states of a hypothetical cohort of patients according to the natural history of the disease from a starting time point until the end of an adequately long time horizon[6,7]. Fig. 1 illustrates six mutually exclusive health states for cancer treatments that are possible in each Markov cycle. Markov cycles are relatively short to ensure that the probability of health state transition remains stable for the entire cycle period.

Having undergone first-line treatment, patients in the first cycle may reside in one of the following health states: CR, PR, SD, PD (and pre-progression), or death. In Fig. 1, three states-CR, PR, and SD-are combined into a single state, which can be named clinical response (clinR), standard practice when modeling for cancers that have very high fatality rates. In certain circumstances, the SD state can be separated from the CR and PR states. For instance, if the treatment is expected to be very effective, the CR should be included in the model separately.

In the second cycle, surviving patients for whom first-line treatment failed may be switched to the second-line treatment before their disease progresses further. Some of these patients who do not receive next-line treatment enter the PD state and some of them die. The rest of the patients return to recursive pre-progression state.

Some patients who respond to first-line treatment or who stay in the SD state enter the relapse state in subsequent cycles, and some remain there until death, as a result either of disease progression or being switched to second-line treatment. The rest of the patients stay in the recursive response or stable state. Those who receive second-line treatment follow a similar pattern of health state transition in the subsequent cycles. For models that do not allow patients to be switched to the next-line treatment, the health state transition, indicated by the dashed line, can be omitted from the model.

**Transitional probability**

In order to ensure the model parameters include health outcomes that are similar to those used in clinical studies, the first step with any model of a potentially terminal disease, like cancer, is to calculate the probability of transitioning to the death state from the comparator arm. The probability of dying, P (D), in the first cycle can be calculated using an OS derived from a single study, over a duration comparable to the cycle length, as shown in equation 1.

\[
P(D) = 1 - OS \tag{1}
\]

If the data were obtained from several studies (k = 1, 2, …, K), an average for the P (D) should be derived using the weighted average, P (D)avg, by placing a higher weight on data from studies with larger sample sizes (n). Because probability (P) is non-linear, pooling the studies need to be done through a transformation and retransformation of the linear parameter rate (r), as shown in equations 2-4[3].

\[
r(D)_k = \ln (1 - P(D)_k) \tag{2}
\]

\[
r(D)_{avg} = \frac{n_1 r(D)_1 + n_2 r(D)_2 + ... + n_k r(D)_k}{n_1 + n_2 + ... + n_k} \tag{3}
\]

\[
P(D)_{avg} = 1 - \exp (-r(D)_{avg}) \tag{4}
\]

For the treatment of interest, estimating P(D) directly from the treatment arm of a study, as shown in equation 1, is not recommended. Instead, the relative efficacy of OS should be used to enable the treatment
to be compared with the control. For example, the treatment’s $P(D)$ should be derived from the hazard ratio (HR), as shown in equation 5. In this case, the HR can be obtained from a meta-analysis as well.

$$P(D)_{treatment} = \text{HR}_{OS} \times P(D)_{control}$$

Estimating the probability of the disease progressing, $P(PD)$, is slightly more complex. It is not exactly equal to $1 - \text{DFS}$ (or $1 - \text{PFS}$), since the DFS or PFS is a combined end point, which includes death or $P(D)$, as shown in equations 6 and 7.

$$\text{DFS} = 1 - P(D) - P(PD)$$

$$P(PD) = \text{OS} - \text{DFS}$$

To distinguish between those patients for whom treatment failed and those for whom it did not, using the treatment comparator, the probability of having a clinical response, $P(\text{clinR})$ for the whole cohort is used. This can be estimated using the objective response rate (ORR), which is based on the data from surviving patients, as reported by clinical studies (equation 8).

$$P(\text{clinR}) = (1 - P(D)) \times \text{ORR}$$

For the treatment of interest, an estimation of the $P(\text{clinR})$ should not be derived directly from the ORR of the study’s treatment arm. Instead, relative efficacy of DFS or PFS, derived from comparison of the treatment with the control should be used for this purpose, as shown in equation 9.

$$P(\text{clinR})_{treatment} = (1/\text{HR}_{DFS}) \times P(\text{clinR})_{control}$$

The probability of a patient’s treatment failing or of a patient being in the pre-progression state, $P(\text{pre PD})$ of whom some would continue onto next-line treatment, is estimated in equation 10.

$$P(\text{pre PD}) = 1 - P(D) - P(\text{clinR})$$

Model calibration

Before using the hypothetical cohort simulation to conduct further calculations on cost and effectiveness, the constructed model should be examined for its validity. This can be done by examining whether the results of the cohort simulation are consistent with findings from the referent clinical studies or the established epidemiology of the disease. Frequently, a number of parameters need to be calibrated again to refine the model, until the simulation result is close to the real patient data.

For cancers, the best indication of a valid model is the survival curve, which most clinical studies use as a primary endpoint. Fig. 2 and 3 compare the survival at various time points between data obtained from the models and data reported by the RCT or national registry. Notice that the survival curves are comparable, even though in Fig. 2 the follow-up period in most RCTs is shorter than the time horizons used in the models.

Guidelines for Health Technology Assessment in Thailand (second edition): Recommendations for economic evaluation of cancer treatment

The authors have presented a summary of
each of the four key aspects that should be taken into account when constructing a Markov model of cancers: 1) choice of outcome measures, 2) health state transition, 3) transitional probabilities, and 4) model calibration. In developing a Markov model for cancer, the first step should be the clarification of the probability of transition to the death state. This can be estimated using the OS, obtained directly from the end point of existing clinical studies. For models that permit next-line treatment, a secondary end point, usually in terms of response to the first-line treatment, can also be included. Validity of the constructed model should then be verified by examining the survival results of the models and comparing them with data from existing databases. The parameters should then be adjusted accordingly, until the model findings are close to reality.

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
การสร้างตัวแบบการเปลี่ยนสถานะสุขภาพสู่การประเมินทางเศรษฐศาสตร์ของการรักษาโรคเรื้อง

จุฬารัตน์ อิมวันเต็ม, สุทธิ อิมวันเต็ม

บทความนี้ย้ายพื้นฐานสำคัญ 4 ประการ ในการสร้างตัวแบบมาร์คด์ฟอร์มหรือรูปแบบที่มีแผ่น  มัลติเวิลด์พัสดุ การเปลี่ยนสถานะสุขภาพ ความจำเป็นของการเปลี่ยนสถานะสุขภาพ และการบริหารแบบ  การเปลี่ยนสถานะสุขภาพเป็นการศึกษา เนื่องจากผลและผลที่เกิดได้จากแนวความจำเป็นของการเปลี่ยนสถานะสุขภาพที่เป็นไปได้ ผลและผลจะกระทบต่อนักการดำเนินการ และแบบแผนการรักษา
ความจำเป็นต้องมีการสร้างตัวแบบที่ส่งข้อมูลสรุปข้อมูลการตอบสนองได้เป็นอย่างดีในระดับของผู้รับผิดชอบด้วยการปรับปรุงการประเมินผลกระทบโดยทั่วไปของกิจกรรมที่เกี่ยวข้องกับการรักษาโรคเรื้อง
Quality Assessment of Health Economic Evaluation

Usa Chaikledkaew BSc(Pharm), MA, PhD*,**, Kankamon Kittrongsiri BSc(Pharm)**

* 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

In Thailand, the results of economic evaluations have increasingly been used to help improve the transparency of health technology prioritization and to inform the resource allocation decision-making process. However, variable quality can mean that application of study results can be limited. To help improve uniformity and widen the application of results, quality assessment of health economic evaluations is crucial. By subjecting health economic evaluations to a rigorous quality assessment process, decision-makers can choose to only use findings from studies that reach the appropriate standard as the basis for policy-making. This article gives a summary of the three key areas to examine when assessing quality - 1) data sources, 2) result reporting and 3) analysis methodology. It is hoped that this will help provide critical guidance to users of economic evaluation results to ensure that they understand and are able to perform quality assessment prior to applying study findings.

Keywords: Quality assessment, Economic evaluation, Thailand

J Med Assoc Thai 2014; 97 (Suppl. 5): S113-S118
Full text. e-Journal: http://www.jmatonline.com

Health economic evaluation (HEE) is one of the approaches used in health technology assessment (HTA). In Thailand, the results of economic evaluations have increasingly been used to help improve the transparency of health technology prioritization and to inform the resource allocation decision-making process(1,2). Subjecting HEEs to quality assessment ensures that studies are of the requisite standard and helps determine whether they are fit to inform policy decision-making. Increasingly, policy-makers are using the evidence garnered from health economic evaluations to inform their policy decisions. This can be seen, for instance, in the drug regulation authority’s drug registration process and the selection process for inclusion of drugs on the National List of Essential Medicines or hospital formulary, both of which rely in part on data from economic evaluations. Clearly then, it is important that health professionals have access to information regarding the quality of the health economic evaluations from which they are taking data.

Why assess the quality of health economic evaluation studies?

Data from health economic evaluations (HEE) should only inform policy decision-making when users understand the three main reasons why assessing the quality of these health economic evaluations is important.

First, quality assessment ensures that only appropriate HEE studies are used to inform policy. This means that only HEE studies that compare the costs and outcomes of at least two health interventions should be used. However, most studies that refer to themselves as a cost-effectiveness analysis do not fulfill these criteria. Any study that either evaluates only costs or does not compare two or more interventions should be regarded as invalid as a basis for policy decisions regarding cost-effectiveness. A diagram of this process is shown in Fig. 1.

Second, quality assessment ensures that the quality of the information generated by the HEE is of the requisite high quality. Teerawattananon et al’s recent study assessing the quality of result reporting in HEE studies within the Thai context found that both the quality and quantity of HEE studies in Thailand was still limited(3). Their study also revealed significant variation in the methods used, meaning that the comparison of data from different studies is very difficult; this is surely due in part to the absence of
economic evaluation guidelines specific to Thailand. HEE information of a high quality should only be regarded as of use to policy decision-makers where it is performed correctly and reported accurately; a lack of high quality HEE studies should be regarded as a barrier to effective policy decision-making (6-12).

Third, quality assessments allow the reliability of the HEE results to be evaluated appropriately. Since HEE studies can be performed using a modeling approach and researchers can input parameter data to predict the cost-effectiveness of a given health intervention, there is a tendency for manipulation of the results to occur, which can lead to unreliable HEE results. Quality assessment of HEE studies limits this and ensures greater result reliability.

A method for selecting health economic evaluation studies for quality assessment

Once HEE studies have been identified, each study should be examined to see if it fits the criteria for quality assessment. Any studies that fail to fulfil the criteria should not be used for quality assessment, although they should still be retained for reference. There are two criterion that need to be fulfilled (Fig. 1): 1) the study must compare at least two choices of interventions and 2) the study must evaluate both costs and outcomes.


Once the appropriate studies have been identified, quality assessment can begin. The quality assessment examines three areas: 1) data sources 2) result reporting and 3) analysis methodology.

Quality assessment of data sources

Economic evaluation studies rely on numerous clinical and cost data sources. The quality of the data garnered from these sources will affect the quality of the analysis; thus, assessment of their quality is crucial. The quality of cost data sources in economic evaluations is assessed according to the hierarchy of evidence (see chapter “Measurement of Costs for Health Economic Evaluation” in this volume) and the quality of clinical data is assessed according to the hierarchy of clinical evidence (see “Measurement of Health Outcomes” in this volume).

Quality assessment of result reporting

The quality of the result reporting is assessed using the criteria developed by Drummond et al (4,5). A summary of these criteria is given below:
- The study perspective is clearly defined.
- The characteristics of the compared intervention are described.
- Discounting for both costs and outcomes of the study period is greater than one year.
- Incremental cost-effectiveness ratio (ICER) is calculated.
- Uncertainty analysis is performed.
- All funding sources are disclosed.

Quality assessment of analysis methodology

A number of guidelines have made recommendations on how best to assess the quality of the methodology used in economic evaluations. These can be divided into two types: guidelines established by the national bodies responsible for performing economic evaluations in each country (e.g. Australia (14), Canada (15), Denmark (16), Norway (17), Hungary (18), England and Wales (19), and Thailand (13)) and guidelines developed by health economists (i.e. Drummond et al (4,5), Gold et al (20), and Tan-Torres (21)).

To ensure that all economic evaluations are transparent, easily comparable, and of high quality, in this second edition of HTA guidelines for Thailand, we outline a specific reporting format for researchers to follow, comprising of ten key elements, all of which should be included. A report checklist has been developed to help guide this process (Table 1). The report checklist can be used alongside the guidelines to assess whether HEE studies can be used by decision-
Table 1. Checklist for quality assessment for health economic evaluation studies

<table>
<thead>
<tr>
<th>Criteria for quality assessment for economic evaluation studies</th>
<th>Answer</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. State the background of the problem</td>
<td>[ ] Yes</td>
<td>[ ] No</td>
</tr>
<tr>
<td>2. State the economic importance of the study</td>
<td>[ ] Yes</td>
<td>[ ] No</td>
</tr>
<tr>
<td>3. State the clinical importance of the study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. State the objective of the study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. State the target population for intervention</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. State the perspective of the study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. State the time horizon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. State the type of economic evaluation methods (i.e., CMA, CBA, CEA, or CUA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. The type of economic evaluation method is appropriate to the study objective</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. State the design of the analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. State the description of all interventions in the analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. State the rational of selecting the comparators in the analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cost and effectiveness data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Identify the outcome measured in the study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. State the sources of effectiveness data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. State the study design of effectiveness data (if one study was used.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. State the description of meta-analysis in synthesizing effectiveness data (if multi-study were used.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. State the valuation of utility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. Summarize effectiveness parameters in table</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Cost data components are in accordance with study perspective</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. State the sources of cost or charge data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. Describe the method of collecting indirect cost and direct non-medical cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22. State the resource use separately from the cost data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. State the valuation of resource use and unit cost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. State the year of valuation for all costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25. State details provided of any adjustment for inflation/deflation for all costs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26. State the currency unit of cost data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. In the case of exchanging money values, state the exchange rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. State the method of transforming charges into costs or costs into charges</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. In the case of using data from expert opinion, state the sources and methods used to collect the data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. In the case of the study period being longer than 1 year, state whether discounting has been performed for costs and/or effect.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Yes* [ ] *No* [ ]

If answers are “Yes” for both questions, please continue to the below checklist

If answer is “No” for any question, the study is not full economic evaluation
### Table 1. cont.

<table>
<thead>
<tr>
<th>Criteria for quality assessment for economic evaluation studies</th>
<th>Answer</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>31. In case of the study period being longer than 1 year, state the discount rate</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>32. State the rationale of using the chosen discount rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. In cases of discounting has not been performed, state the rationale</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. Summarize cost parameters in table</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>35. Describe the event pathway in the model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36. Show a diagram of event pathways in the model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37. State the software used in the model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38. State the details of model validation that have been provided</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39. State the time horizon used in the model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40. For Markov models, state the cycle length of the model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. State all assumptions used in the model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42. Perform the sensitivity analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43. State the sensitivity analysis method</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44. State the choice of variables and the ranges used in the sensitivity analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45. Describe the rationale of chosen parameters in the sensitivity analysis</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>46. When performing the incremental analysis, all relevant interventions are included.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47. Report the results of the incremental analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48. Present the results of the undiscounted reference case values i.e., total cost, total effectiveness, incremental cost, incremental effectiveness, and incremental cost-effectiveness ratio</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>49. Explain the summary of the reference case results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50. Present the important disaggregated and aggregated results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51. Present the result in graph i.e. the cost-effectiveness plane</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52. Present the sensitivity analysis i.e. tornado diagram or cost-effectiveness acceptability curve</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>53. State the conclusion of sensitivity analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54. In case of the budget impact analysis performed, state the analysis result</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>55. State the answers for research questions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56. State the conclusion in accordance with the reported results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57. State the conclusions and appropriate precaution of the study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58. Explain the feasibility of the application of study results on policy decision making</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>59. Discuss the important ethical implications</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60. Explain the limitations of the study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61. Compare the results with other studies' results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62. State the impact on annual budget</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63. State the funding sources for the study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64. State the author's conflict of interest with the funding sources</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1Answer “Yes” if the study clearly specify according to criterion
Answer “No” if the study not clearly specify according to criterion
Answer “Not applicable” if the criterion not applicable to the analysis i.e. The criteria of modeling approach are not applicable for non-model study.
When answer “No”, description of missing or irrelevant data should be specified.
makers as a basis for policy. The checklist is not limited to administrators and policy-makers at the national and local level. Indeed, if stakeholders such as pharmaceutical companies were required to submit HEE information about their products alongside the clinical information which they already have to supply, they could use the checklist as a key tool to generate data that might be very beneficial to those who decide which technologies are included in health benefit packages. However, the checklist should only be regarded as a tool to help guide the preliminary phase of quality assessment; it is not intended to be used to judge the quality of study’s methodology or results.

Method for scoring the quality of economic evaluation studies

Many studies have used the checklist to allocate a score to each question, after which the individual scores are added together to arrive at a total final score that is indicative of a study’s overall quality. A review of the existing literature identified six studies that outlined suggested scoring systems for the quality assessment of health economic evaluations. However, no uniform scoring approach was found that is both valid and reliable for the quality assessment of methodology used in economic evaluation. As such, the use of a scoring system for the quality assessment of methodology used is not recommended\(^{22}\). Instead, results from the quality assessment should be presented in the same format as the checklist, with a description of the results and how they compare to the criteria presented alongside. Moreover, full explanation of the methodology used and the results should be given, along with a description of the key strengths and weaknesses of the present study that may affect the reliability of the results.

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

3. Teerawattananon Y, Russell S, Mugford M. A systematic review of economic evaluation literature in Thailand: are the data good enough to be used by policy-makers? Pharmacoeconomics 2007; 25: 467-79.
12. Barbieri M, Drummond M, Willke R, Chancellor J,


Application of HTA Research on Policy Decision-Making

Sitaporn Youngkong MSc, PhD*

* Health Intervention and Technology Assessment Program, Ministry of Public Health, Nonthaburi, Thailand

This article provides an overview of the potential uses of health technology assessment (HTA) in health technology or health intervention-related policy decision-making. It summarises the role of HTA in policy planning, health system investment, price negotiation, development of clinical practice guidelines, and communication with health professionals. While the multifaceted nature of HTA means that some aspects of the data can result in conflicting conclusions, the comprehensive approach of HTA is still recommended. To help minimise the potential conflicts within HTA data, a multi-criteria decision analysis (MCDA) approach is recommended as a way to assess a number of decision criteria simultaneously. A combination of HTA with MCDA allows policy decision-making to be undertaken in an empirically rigorous and rational way. This combination can be used to support policy decision-makers in Thailand and help them prioritise topics for assessment and make informed health benefit package coverage decisions. This approach enhances the legitimacy of policy decisions by increasing the transparency, systematic nature, and inclusiveness of the process.

Keywords: Health Technology Assessment, Multiple-criteria Decision Analysis, Policy decision-making

J Med Assoc Thai 2014; 97 (Suppl. 5): S119-S126
Full text. e-Journal: http://www.jmatonline.com

In Thailand, as in many countries nowadays, the availability of high-cost health interventions, including pharmaceuticals and medical technologies, is on the rise. While this has frequently led to an increase in public and patient expectations, the resource-limited nature of government healthcare systems mean that these interventions cannot always be provided for those who wish to have them. There is currently, therefore, a need for rigorous, efficient, and evidence-based approach to help policy-makers decide allocation of limited healthcare budgets. Health Technology Assessments (HTAs) are increasingly recognised as one of the most useful tools that can help inform health technology- or health intervention-related policymaking at individual, institutional, national, and international levels(1,2). HTA data can improve the quality of health care in multiple ways, and is a valuable resource for the development of a wide spectrum of health care policies(3). This section provides an overview of the potential applications of HTA for policy-makers when making decisions on the most rational use of health.

Policy planning and investment in the health system

HTA can be used in policy planning to help determine an appropriate level of health care investment, given the resources and health technologies that are available. HTA can provide information on the budget impact of health technology adoption within a specific country setting, as well as support for decision-makers regarding resource allocation of different health initiatives, including those used in the selection of health interventions for public reimbursement. For instance, HTAs can support decision-makers in deciding which benefit packages they will provide, as shown in case study 1.

Case study 1: Inclusion of allogeneic hematopoietic stem cell transplantation for severely thalassemic patients in the Universal Health Coverage Benefit Package in Thailand(4)

Hematopoietic stem cell transplantation (HSCT) is the only curative treatment currently available to leukaemia and severely thalassemic patients. At the time of the research, HSCT was not included in the benefit package of the Universal Coverage (UC) scheme but was covered in the Civil Servant Medical Benefit Scheme and the Social Security Scheme. This resulted in inconsistent coverage for the population. Of those patients whose insurance did not cover HSCT provision, only those who could afford to pay for the treatment and the associated expenses (around 700,000-
1,500,000 THB per patient) were able to access the treatment. The National Health Security Office (NHSO), which manages the UC scheme, examined a number of health problems/interventions and, through a rigorous selection process facilitated by the Health Intervention and Technology Assessment Program (HITAP) in 2007, identified five that warranted further economic evaluation to determine whether the coverage policy related to each was appropriate. Once the top health problems/interventions had been identified, the Subcommittee for the Development of the Benefit Package and Service Delivery (SCBP) of the NHSO approved the selection. Then, an economic evaluation and budget impact analysis were conducted by HITAP researchers. Of the five health topics/interventions, the intervention that was deemed the top priority for further assessment proposed examining patients with severe thalassemia and their caregivers who were undergoing HSCT.

A cost-utility analysis was performed to evaluate and compare the costs and health outcomes of HSCT compared with standard therapy—in this case, iron chelating therapy (ICT). The results showed that the incremental cost-effectiveness ratios (ICERs) related to HSCT (for patients aged 1 to 15 years) ranged between 80,700 and 183,000 THB per QALY gained, while those associated with the current treatment package (for patients aged 1 to 15 years) ranged from 209,000 and 953,000 THB per QALY gained. Based on a willingness to pay threshold of 100,000 THB per QALY (approximating the Thai GDP per capita), the findings of the analysis indicated that provision of HSCT to severely thalassemic patients with related or sibling donors were likely to be cost-effective only when provided to patients 10 years and under. Moreover, the governmental budget impact of providing HSCT to patients 10 years and under equated to around 90 million THB per year (with only 200 patients requiring treatment per year).

The results of the assessment were subsequently presented to the SCBP for appraisal. The SCBP agreed to provide HSCT coverage for patients aged 10 years and under. This decision was not only based on the fact that the treatment was found to be cost-effective, but also because the subcommittee deemed it socially, ethically, and morally appropriate to do so. However, before including this technology in the UC benefit package, the SCBP asked the researchers to conduct further research on the feasibility of providing this life-saving and cost-effective technology on a large scale in an equitable manner, to examine further whether limited resources should be allocated based on certain criteria, e.g. severity of disease, fair innings, or a lottery model.

**Price negotiation**

In recent years, HTA data (mostly using ICER information) have been used in drug pricing negotiation. In HTA studies, the price that is used in the analysis is usually the maximum amount that the pharmaceutical company hopes to charge. If the drug demonstrates good value for money the company may be awarded a price similar to that assumed in the assessment. However, the drug price may be negotiated downward based on the results of the economic evaluation and other relevant information presented in HTA. An example of this can be seen in case study 2.

**Case study 2: Using the results of a cost-utility analysis of drug treatments in patients with chronic Hepatitis B for drug price negotiation**

An estimated 2 to 3 million people in Thailand are chronically infected with the Hepatitis B virus. Infection with Chronic Hepatitis B (CHB) can cause not only liver inflammation and serious liver damage leading to cirrhosis, hepatic decompensation, hepatocellular carcinoma and death, but may also result in a significant economic burden for patient and caregiver. Six medications—lamivudine, adefovir, entecavir, telbivudine, interferon and pegylated interferon—have been licensed by the Thai Food and Drug Administration (Thai FDA) for the treatment of CHB; only lamivudine has been included in the National List of Essential Medicines (NLEM). To date, no intervention to help manage drug resistance in CHB patients has yet been included in the NLEM. However, tenofovir, approved by the Thai FDA for treating HIV, is currently used for CHB treatment in clinical practice, since it has demonstrated high antiviral efficacy and low rates of resistance in CHB patients.

To assess whether this clinical practice was, indeed, efficient and to help decide which drug should be included in the NLEM, a cost-utility analysis of all treatment options for patients with CHB in Thailand was conducted. The study found that tenofovir was cost-effective when used as a first-line treatment (or second-line treatment in the case of lamivudine resistance), and as such, should be included in the NLEM. However, the inclusion of tenofovir in the NLEM was also found to potentially have a high budget impact for the government, based on the market price (2010). Therefore, the subcommittee of development of the...
NLEM negotiated with the pharmaceutical companies to reduce the price of tenofovir. They reduced the price from 43 baht to 12 baht per tablet, resulting in a saving of 375 million baht of government money.

**Development of clinical practice guidelines**

HTA is one of the most useful tools available when developing clinical practice guidelines. In some developed countries, such as England and Wales, HTA results are always taken into account when developing any guidelines intended to influence health service delivery throughout the country.

**Communicating with health professionals**

It is also widely agreed that HTA is a useful resource for public health authorities involved in communicating and promoting rational health technology use to health professionals.

A cornerstone of many national disease programs is the capacity of interventions to maximise general population health, otherwise known as effectiveness. Some national disease programs also propose that cost and cost-effectiveness are also important criteria that should be used to guide choices in health care. These criteria explicitly recognise the economic constraints on the provision of health care, and imply that only those interventions that show value for money should be publicly financed. However, over reliance on these kinds of criteria places disproportionate emphasis on the economic aspects of treatment and fails to capture other important aspects of health care that should be taken into account when evaluating interventions. HTAs are more holistic, taking into account different types of evidence, including that related to safety, efficacy, economics (value for money and budget impact), feasibility, societal, and ethical implications of implementing health technologies or interventions within the health system. Because of HTA’s multifaceted nature, at times, these different dimensions may conflict with one another. While the HTA model may be more complex that a purely economic evaluation of the intervention, this more comprehensive approach does allow decision-makers to set priorities and allocate limited resources among different health programs in the most informed way possible. However, many HTA organisations (such as the National Institute for Health and Clinical Excellence (NICE) in England and Wales) have raised concerns about how best to address all of these issues simultaneously; currently, there is no consensus on how best to resolve this issue.

Researchers have suggested that a multi-criteria decision analysis (MCDA) approach, used together with an HTA, might be more appropriate for ensuring that analyses incorporate all relevant issues into their framework, not just effectiveness and cost-effectiveness, to better inform the policy making process.

**Multi-criteria decision analysis**

A multi-criteria decision analysis (MCDA) is a decision support method that allows the identification of a comprehensive set of criteria, establishes the performance of interventions on those criteria in a so-called performance matrix, and then inspects the performance matrix qualitatively or quantitatively to rank order interventions. In a qualitative inspection, policy-makers simply interpret the performance matrix, and make implicit judgments on the weights of the various criteria. In a quantitative inspection, policy-makers weight the different criteria on the basis of their relative importance, and multiply the score by the weights to obtain weighted averages for all interventions. Interventions can subsequently be ranked ordered according to these weighted averages. While MCDA reduces the stream of dissimilar information by assessing the interventions’ performance according to a set of criteria in the performance matrix, the consideration of other non-quantifiable (or otherwise absent) criteria that did not present in the performance matrix for any reason is captured in the process of deliberation. Decisions on health intervention priorities should be made by using MCDA, on the basis of consultations with the relevant multiple stakeholders through a deliberative process. Therefore, a combination of quantitative and qualitative approaches is recommended.

In recent decades, the use of an MCDA approach in HTA has grown in popularity, as a way of improving the evidence upon which healthcare decision-makers base their healthcare policy decisions at all levels, and to ensure better synthesis, communication, and dissemination of HTA results. Combining MCDA and HTA has been found to provide higher quality data that encourages more rigorous policy planning in the long run and allows policymakers to make informed decisions that are relevant to their specific context, either at the a national, sub-national or institutional level. In Thailand, the lessons learned from the two case studies presented here in provide ample evidence that MCDA has been useful both as a way to select health topics that warrant further HTA studies, and as a tool to support health
technology coverage decisions in the Thai UC benefit package\textsuperscript{(5,20,21)}.

Due to limited resources, many HTA agencies struggle to keep pace with new technology. As such, priority setting has become a crucial aspect of the process to determine which health products/topics are assessed. Identification of technologies that warrant assessment is now the first step of the HTA process, and it helps ensure that HTAs continue to be relevant and conducted in a transparent, systematic, and socially acceptable manner\textsuperscript{(22)}. Clearly, priority setting is an important process that helps makes sure that HTA offers optimal benefits for society and encourages the use of HTA data in policy decision-making. However, to date, very little research has been conducted into the criteria that should be applied and the process that should be undertaken when selecting HTA topics. Nevertheless, there is some agreement on general topics that should be examined (Table 1) as well as growing acceptance, based on examples, that MCDA can help inform this decision (case study 3).

Case study 3: Selection of health topics for health technology assessment in Thailand, facilitated by HITAP

In 2007, HITAP initiated some research on appropriate methods for identifying priority health topics/technologies that warrant HTA. To ensure that all decision-making resulting from HTA data is systematic, transparent, and efficient\textsuperscript{(24-26)}, the MCDA approach was applied to the priority setting process (Fig. 2).

Step 1. Nomination of health topics/technologies for assessment

As the first step of the process, HITAP invited potential HTA users to nominate a number of health topics/technologies that they considered warranted further assessment. The potential HTA users included health professional councils, health care purchasers (i.e., the National Health Security Office, the Social Security Office, and the Comptroller General’s Department), public health agencies (at provincial and national levels), academia, private sector, civil societies, patient groups, and the general population.

Step 2. Short listing and Literature Review

In the second step, HITAP researchers shortened the list of nominated health topics/technologies by excluding all of the following: 1) those that had been assessed within the last five years; 2) those whose assessment was not the responsibility of HITAP but, rather, an officially authorised organisation;

![Fig. 1](image1.png) Ad hoc priority setting and rational priority setting\textsuperscript{(17)}.

![Fig. 2](image2.png) A schematic representation of HTA topic selection process, facilitated by HITAP.
### Table 1. Examples of potential criteria used for topic selection of HTA(22), by category

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease burden</td>
<td>High individual burden of morbidity, mortality, or disability</td>
</tr>
<tr>
<td></td>
<td>High population burden of morbidity, mortality, or disability</td>
</tr>
<tr>
<td></td>
<td>Prevalence or incidence</td>
</tr>
<tr>
<td></td>
<td>Disease-adjusted life expectancy or healthy years of life expectancy</td>
</tr>
<tr>
<td>Clinical impact</td>
<td>Potential health impact of the proposed technology versus standard care in a person with a clinical condition</td>
</tr>
<tr>
<td></td>
<td>Benefits of conducting an assessment in terms of reduced clinical uncertainty</td>
</tr>
<tr>
<td></td>
<td>Potential for change in practice to affect patient clinical outcomes</td>
</tr>
<tr>
<td>Budget impact</td>
<td>High aggregate cost of a technology or health problem</td>
</tr>
<tr>
<td></td>
<td>Potential incremental budgetary impact of adoption of the technology in comparison to the current standard of care</td>
</tr>
<tr>
<td>Economic impact</td>
<td>High unit healthcare cost of a technology or health problem</td>
</tr>
<tr>
<td></td>
<td>Potential cost-effectiveness of conducting an assessment given resource constraints (e.g., researchers, time, and research funding) of the assessment program</td>
</tr>
<tr>
<td>Variation in practice</td>
<td>Variation in rates of use of the proposed technology for the given clinical condition</td>
</tr>
<tr>
<td></td>
<td>Substantial variations in clinical practice</td>
</tr>
<tr>
<td>Expected level of interest</td>
<td>Public or political demand</td>
</tr>
<tr>
<td></td>
<td>Scientific controversy or great interest among health professionals</td>
</tr>
<tr>
<td></td>
<td>Need to make regulatory decision</td>
</tr>
<tr>
<td></td>
<td>Need to make a health program implementation decision</td>
</tr>
<tr>
<td></td>
<td>Need to make payment decision</td>
</tr>
<tr>
<td></td>
<td>Social and ethical implications associated with the use of the proposed technology</td>
</tr>
<tr>
<td>Evidence</td>
<td>Sufficient research findings available upon which to base assessment</td>
</tr>
<tr>
<td></td>
<td>Timing of assessment relative to available evidence</td>
</tr>
<tr>
<td></td>
<td>Potential for the findings of an assessment to be adopted in practice</td>
</tr>
</tbody>
</table>

3) those whose performance was already well known; and 4) those that were not directly related to health.

Once the inappropriate health topics/technologies were dismissed, each short-listed topic/technology was reviewed and scored according to a set of criteria, developed by HITAP on the basis of a review of international guidelines used to prioritise HTA topics for assessment (as presented in Table 1). The list of criteria was presented to a number of HTA experts in Thailand, who refined the list to five criteria: 1) disease burden, 2) economic impact on household expenditure, 3) variation in practice, 4) potential for the findings of an assessment to be used, and 5) equity/ethical and social implications (Table 2). The shortlisted topics/technologies were then ranked according to how they scored according to this list of criteria.

### Step 3. Topic selection for assessment

Decision-making is a complex process, and resource allocation is always inherently political. Although a comprehensive set of criteria have been developed in terms of how best to undertake the priority setting process, decision-makers should still be permitted room to elaborate their own reasons in the final step, to ensure the process is deliberative. To this end, the shortlisted health topics/technologies were presented to multiple stakeholders in a consultative meeting. The participants were asked to discuss which topics should be assessed by HITAP, and present a final list.

To conclude, HTA is one of the most useful tools for decision-makers involved in healthcare policy. It can help prioritize topics for assessment, and inform coverage decisions of health benefit packages. In addition, the adoption of an MCDA approach when conducting HTAs can help reduce the stream of dissimilar information allowing assessment of intervention performance. The combination of HTA and MCDA can help the policy planning process in the long term and enhance the legitimacy of
potential decisions by increasing the transparency, inclusiveness, and accountability of the process.

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, 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


การใช้ประโยชน์จากงานวิจัยตามการประเมินเทคโนโลยีด้านสุขภาพ

คิตตินร วงศ์ไชย

บทวิเคราะห์นั้นเสนอประสบการณ์การใช้ประโยชน์จากการประเมินเทคโนโลยีด้านสุขภาพ สำหรับการคัดเลือกใช้เทคโนโลยีที่มีประสิทธิภาพ หรือการคัดเลือกสุขภาพที่เกี่ยวข้อง ทำให้การคัดเลือกใช้เทคโนโลยีด้านสุขภาพ ในการวางแผนบริการสุขภาพ การจัดระบบบริการสุขภาพ การจัดทำแนวทางการปฏิบัติ และการเปลี่ยนแปลงการบริการสุขภาพ รวมทั้งการประเมินเทคโนโลยี ด้านสุขภาพได้หลากหลาย และในผลดังกล่าวได้สรุปได้ดังนี้ คือ การเปรียบเทียบปรากฏการณ์ที่มีผลกระทบต่อการใช้เทคโนโลยีด้านสุขภาพ หรือการเปลี่ยนแปลงการคัดเลือก สิ่งที่ได้สืบเนื่องจากการวิจัยที่ได้กล่าวกับ การวางแผนบริการสุขภาพ การจัดทำแนวทางการปฏิบัติ และการเปลี่ยนแปลงการบริการสุขภาพ หรือการเปลี่ยนแปลงการคัดเลือก สิ่งที่ได้สืบเนื่องจากการวิจัยที่ได้กล่าวนั้น การที่จะทำให้การคัดเลือกสิ่งที่ได้สืบเนื่องจากการวิจัยเป็นกระบวนการที่มีประสิทธิภาพ และพิจารณาอย่างรอบคอบ

Standard Cost Lists for Health Economic Evaluation in Thailand

Arthorn Riewpaiboon BPharm, MSc, PhD*

* Social and Administrative Pharmacy Excellence Research (SAPER) Unit, Department of Pharmacy, Faculty of Pharmacy, Mahidol University, Bangkok, Thailand

This analysis was undertaken to generate a set of standard costs for medical services and those incurred by patients receiving treatment, for use in health economic evaluations. Medical service unit cost data were derived from a survey of 3,091 hospital medical services in five hospitals, disaggregated by type of hospital (district or provincial/regional) and analyzed using the relative value unit method. Patient-borne ambulatory cost values were derived from data gathered through 905 patient interviews that took place in six health centers, three district hospitals, and three provincial/regional hospitals. The survey gathered data on costs arising from the distance travelled to access the medical service, the time spent in the healthcare facility, as well as travel and meal costs. The analysis generated a set of standard cost data for Thailand that will make conducting economic evaluations more accurate, faster, and more convenient, as well as allowing better comparability between studies. This is the first standard cost menu that has been developed specifically for Thailand, and as such should be revised and refined in the future. Some areas that would benefit from revision are suggested.

Keywords: Medical service, Unit cost, Standard cost list, Health economic evaluation, Thailand

Economic evaluation is a tool that is widely used tool to aid decision-making, both from a practical, individual standpoint and as part of the development of national policies. Despite this, implementation of the approach is not always straightforward. One issue that can impede accurate economic evaluation is the process of obtaining accurate cost measurements, which may be derived using a variety of concepts, methods, and reference values(1,2). This can lead to instances where one technology is assigned various different values, as a result of the researchers using different methods and/or references in the calculations. This difference in costing may not necessarily reflect a real difference in resource usage, but merely a different calculation approach on the part of the researcher. One recent study in Thailand found that the capital cost of a district hospital calculated using an economic approach was 13% higher than that calculated using an accounting approach. The same study also found that using a 6% discount rate comparing to 3% rate increased the calculated cost by 4.8%(3).

The costing process involves three basic steps: identification, measurement, and valuation of resource use(4,5). While the first two steps are relatively straightforward, the final step—valuation—involves multiplying the unit cost of the resource in question by the quantity used. For medical services, the unit cost can be determined by direct measurement, standard or reference lists for both price and cost, average market prices, or estimation(6). While the specific objective and limitations of some studies may mean that average market prices, or estimation are used to determine unit cost, in general, the use of direct measurement or standard/reference cost lists is preferred. Using direct measurement at a study site to ascertain unit cost is appropriate when the results are to be used for organizational management, while standard unit cost is used when the results are needed for national-level management.

In Thailand, health economic evaluation (HEE) is a standard tool that is used to inform policy development. For instance, before a new drug is included on the national list of essential drugs it must first be subjected to an HEE; the same is true for a new treatment regimen, before it is accepted for coverage under the national health insurance benefit package. To establish standardized study methods in 2008, the Health Intervention and Technology Assessment Program (HITAP) of the Ministry of Public Health
developed a set of National Health Technology Assessment Guidelines, published in book form and in the Journal of the Medical Association of Thailand (volume 91, supplement 2)(7). These guidelines included a chapter on measurement of costs written by this author(8,9). Although the establishment of national standard guidelines on cost measurement has provided researchers with much more clarity, conducting comparisons between studies at a national level remains a challenge, as researchers continue to use unit costs that cannot be compared. To help address this challenge and provide clear guidance on unit costing processes, HITAP assigned this author to develop a list of standard unit costs of medical services for Thailand. The list is intended to increase the efficiency of study implementation, improve the reliability of data, and allow more accurate cross-study comparison.

Material and Method

Overall methods

The present study employed a standard or conventional costing method, which is comprised of five steps(3): cost centre identification, direct cost determination, indirect cost allocation, full cost calculation, and unit cost calculation. To classify cost centers, it is necessary to know which service outputs and resources are used by those units: transient cost centers provide support to patient service units and absorbing cost centers provide medical services to patients. Direct cost determination is a method used to calculate those costs that are directly incurred by the cost centers-labor cost, material cost, and capital cost. Costs incurred by transient cost centers are allocated to absorbing cost centers. Many alternative allocation methods are available, including direct allocation, step-down allocation, double distribution, and simultaneous equation methods(10), the latter of which is the most complicated but also the most accurate. Full cost is calculated by adding together all direct and indirect costs. Finally, the full cost is allocated to output services (or cost objects or cost products).

There are several unit cost allocation methods. The average method is used for cost centers producing only one service or a set of homogeneous services assumed to be the same service. For multi-service cost centers, there are a number of alternative methods, including micro-costing(11,12), ratio of costs to charges (RCC)(12,13), and relative value unit (RVU) or weight procedure method(12,13).

The micro-costing method is the most accurate since it is based on the real resource consumption of each service. The method starts by measuring the direct cost (labor, material, and capital costs) of each service. After that, the sum of direct costs of all services is subtracted from the full cost, resulting in indirect cost (of the service outputs). Finally, the indirect cost is allocated to each service, and then added to the direct cost to obtain the unit cost.

For the RCC method, the unit price of each service is multiplied by the number of service outputs, which then generates a total charge for that service. After this, the full cost is divided by sum of the charges of all services to obtain the ratio of cost to charge. Finally, this ratio is used to multiply each unit price, resulting in unit cost.

The RVU method is based on the ratio of resources used for all services in terms of standard RVUs. Although this method is not as accurate as the micro-costing method, it offers greater time savings(14,15). First, standard RVUs of all services are developed. Then, the total RVUs used by the hospital are calculated by multiplying the results of the standard RVU by the number of services for all medical services. After that, the cost per RVU is calculated by dividing the full cost by the total RVUs of the hospital. Finally, the cost per RVU is multiplied by the number of RVUs for each service to obtain a unit cost. Standard RVUs can be developed using a ranking method or an objective data method(16-18). The ranking method is a subjective technique that compares resource usage by establishing the smallest amount, and then estimating subsequent amounts in multiples of this initial amount. The objective data method is based on real resource consumption, based either on the consumption of a major selected resource (for instance time or material use), or the costing data derived from other studies.

Specific methods

The standard cost list used here in was developed based on the following sub-research projects conducted by the author, and on theses of graduate students supervised by the author:

1. Development of standard relative value units of health services(19).
2. Unit cost analysis of hospital medical services(20).
3. Direct non-medical costs for outpatients(21).

Development of standard relative value units of health services(29)

The present study was conducted in 2009 to develop standard RVUs for Thailand. The objective...
data method was used for the analysis, based on existing service cost or price lists. In Thailand, there are three main lists—the price list of medical services of hospitals under the Ministry of Public Health(22), which is developed based on costing concepts; the reimbursement list for medical services of public health facilities under the Civil Servant Medical Benefit Scheme (CSMBS)(23), a modified version of the Ministry of Public Health list using assigned service codes available in hospital databases and used by all public hospitals; and the reimbursement list of medical services for road traffic injuries under the victims compensation fund, and the reimbursement list of medical services for road in hospital databases and used by all public hospitals; the Public Health list using assigned service codes available.

**Unit cost analysis of hospital medical services**

To calculate unit costs of hospital medical services, the standard costing approach was used(3). For unit cost calculation, the RVU method(12,13) was employed, using the aforementioned standard RVUs of Thailand(19). Costs were presented based on 2009 values. The study covered regional (>500 beds), provincial (120-500 beds), and district (10-120 beds) hospitals. In larger provinces, regional hospitals provide the same services as provincial hospitals, in addition to offering more advanced treatment. Therefore, regional and provincial hospitals were classified into the same group for the purposes of this study. Only those hospitals that met specific efficiency criteria were included(25). Study sites were composed of three regional/provincial hospitals and two district hospitals. Total hospital costs were calculated, including labor, material, and capital costs but excluding pharmacy costs. Capital costs included cost of using durable assets and opportunity cost of land used. Capital cost of durable assets was calculated using an economic-based approach(10) with a 3% discount rate(30). Useful years were defined according to the guidelines of the Ministry of Finance(26). Items used beyond their useful years were still included in the cost(6,27). After determining total hospital costs, total RVUs were calculated by multiplying the RVU of each service by the total number of service outputs. Then, the cost per RVU was calculated by dividing the hospital’s total cost by the total RVUs. Finally, the cost per RVU was multiplied by the number of standard RVUs of each service, which results in the unit cost.

**Direct non-medical costs for out-patients**

The present study estimated the direct non-medical costs for outpatient services received at health centers, district hospitals, and regional/provincial hospitals by way of a descriptive study using a face-to-face interview technique. Study sites were selected from the central, northeastern, and northern regions of Thailand. In each region, one regional or provincial hospital, one district hospital, and two health centers were selected based on convenience sampling. In each study hospital, approximately 100 patients were selected for interview. For each health centre, approximately 50 patients were interviewed. All study sample patients were at least 18 years old. Patients who came for general physical examinations, appointments for injections, and wound dressing were excluded. The interviews were conducted between October and December 2009, and focused on gathering information on the distance traveled between the home and health facility, time spent, costs involved with transportation, meal costs, and income loss in the course of obtaining medical services. The Mahidol University institutional review board approved the study. Interviewers explained the process of the study to respondents and obtained their written informed consent before conducting the interview.

**Results**

The results from this study generated the first set of standard cost lists for Thailand. These have been published, along with the methodology used, in Thai, in hard copy and online (http://www.hitap.net/research), and in software form (http://www.hitap.net/costingmenu/). Five hundred copies of the book were distributed to academics and related organizations(29).

All costs are given in 2009 values, but these can be adjusted by applying the consumer price index for medical care(29). For international readers, the exchange rate was 34.34 Thai baht (THB) per $1US in 2009(30). The standard cost list is composed of,

- standard RVUs of medical services
- unit cost of medical services at regional/provincial hospitals
- unit cost of medical services of district hospitals
- direct non-medical cost of outpatients at all levels of health facilities

Lists of medical services, standard RVUs, and unit costs at regional/provincial hospitals and district hospitals are presented in Table 1. In the analysis, no data of variability (standard error) is included, as the hospitals in the survey did not provide the same set of services. To broaden the information base, data for
services rendered by all hospitals at the same level were incorporated into one tabulation. The services were composed of 3,091 items in 12 groups, as follows:

- **Group 1:** Routine service at outpatient and inpatient departments (visit and hospitalization day).
- **Group 2:** Blood transfusion services.
- **Group 3:** Diagnostic and clinical pathology services.
- **Group 4:** Diagnostic and therapeutic radiology services.
- **Group 5:** Special investigations.
- **Group 6:** Medical supplies and services.
- **Group 7:** Medical procedures and anesthesia.
- **Group 8:** Nursing care services.
- **Group 9:** Dentistry services.
- **Group 10:** Physical therapy and medical rehabilitation.
- **Group 11:** Acupuncture and other alternative medicine.
- **Group 12:** Health promotion and disease prevention and control.

Some services in the list were not provided at the time of study; in these cases, the unit costs were listed as not available (N/A). The unit cost of these services can be calculated by multiplying the number of RVUs per service (taken from the standard RVU value list) by cost per RVU. The cost per RVU for regional/provincial hospitals and district hospitals was found to be 134.95 THB and 128.67 THB, respectively (2009 values). In the case of services beyond the scope of the list, unit costs were estimated by multiplying the unit price by the cost to charge ratio developed by this program, giving ratios of 1.63 and 1.45 for regional/provincial hospitals and district hospitals, respectively.

To calculate the direct nonmedical cost data for outpatients, 905 patients were interviewed. The interview asked participants about the various factors that contribute to the direct medical costs, specifically—distance travelled from home to health facilities, time spent travelling and receiving services, costs of travel and meals, and real income loss of patients and accompanying persons (Table 2). All values were calculated according to one hospital visit. All data were disaggregated for health centers, district hospitals, regional/provincial hospitals, and the average across all facilities was calculated. To calculate real income loss, the statistical analysis also included persons who had incurred no income loss, due to being on a fixed monthly salary, being self-employed, or being unemployed. In addition, the opportunity cost could be calculated based on the time spent versus a reference wage rate.

**Discussion**

A number of countries have developed standard cost lists to help standardize their economic evaluations; the most well known are those of Australia(31), Canada(32,33), the Netherlands(34) and the United Kingdom(35). The first cost list in health care to be produced was the “Manual of Resource Items and Their Associated Costs”(31). First developed in Australia in 1993 by the Commonwealth Department of Health and Ageing(31), the list gives a standard list of service costs that can be used in economic analyses, the results of which are then submitted to the Pharmaceutical Benefits Advisory Committee. In Canada, the first cost list was developed in Alberta province in 1997; in 1999, a specific cost list for Manitoba health services was developed(36). This list was then incorporated into the national list of provincial costs for health care in 2000(37). In the UK, the first costing guidelines, known as NHS Costing Manual, was first developed in 1998 by the Department of Health. It is revised every year, published in manual form, and provided to all hospitals(36), who then conduct cost analyses based on the values within. Participating hospitals calculate the unit costs of the medical services they provide, and the reference cost list is then developed, based on average costs obtained from data submitted by participating hospitals. In the Netherlands, the first “Dutch Manual for Costing: Methods and Reference Prices for Economic Evaluations in Healthcare” was first published in 2000, and a new and revised version was published in 2010, according to the guidelines on pharmacoeconomic evaluation issued by the Dutch Health Insurance Board. The guidelines have been approved by the Ministry of Health, Welfare, and Sport(34,38,39).

By comparing the Thailand list to other international lists in terms of costing methods, it is clear that the present study used similar costing steps to those conducted in other countries—resource identification, quantity measurement, and valuation of resources used. The main difference is that, while most other countries with costing lists have had regularly revised standard cost lists for more than a decade, in Thailand, this is the first version. A clear benefit of this list is that the results were determined from the calculations using data from actual health facilities meeting criteria of efficiency and quality. Nevertheless, as with all first versions, there are some limitations. The CSMBS reimbursement rate used for the standard RVU development was established several years ago,
which may mean that some of the methods may be slightly out of date. This rate was modified from the Ministry of Public Health price list—a list which was developed by different working groups for different service groups. Each group might therefore have used different methods in determining specific details. Another limitation is the small sample size of health facilities used for the calculation. Future revisions to the list should ideally be performed every few years, and the medical services and corresponding codes should be standardized among the various health facilities. The authors recommend that an institute be established to oversee this job as a continuing responsibility.

**Conclusion**

This is the first standard cost menu to be developed for Thailand. It covers a range of medical services, and covers district hospitals and provincial/regional hospitals. At present, the list does not include services at a super tertiary level or at a university hospital level. This standard cost menu should make

---

**Table 1. Sample of standard RVUs and unit costs of hospital medical services in Thai baht (THB), 2009 values**

<table>
<thead>
<tr>
<th>Service</th>
<th>Unit Code</th>
<th>RVU</th>
<th>Unit cost (THB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 Blood transfusion services</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2 Diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1 Diagnosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1.1 Antibody identification (tube method)</td>
<td>Test 22101</td>
<td>2.0</td>
<td>270</td>
</tr>
<tr>
<td>2.1.2 Antibody identification (gel test)</td>
<td>Test 22102</td>
<td>4.0</td>
<td>540</td>
</tr>
<tr>
<td>2.1.3 Antibody screening, indirect antiglobulin (tube method)</td>
<td>Test 22103</td>
<td>0.5</td>
<td>n/a 64</td>
</tr>
<tr>
<td>2.1.4 Antibody screening, indirect antiglobulin (gel test)</td>
<td>Test 22104</td>
<td>1.0</td>
<td>135 129</td>
</tr>
<tr>
<td>2.1.5 Blood group (ABO) (tube method)</td>
<td>Test 22105</td>
<td>1.0</td>
<td>135 129</td>
</tr>
<tr>
<td>2.1.6 ABO cell grouping</td>
<td>Test 22106</td>
<td>0.5</td>
<td>67 64</td>
</tr>
<tr>
<td>2.1.7 ABO serum grouping</td>
<td>Test 22107</td>
<td>0.5</td>
<td>67 64</td>
</tr>
<tr>
<td>2.1.8 Rh(D) typing</td>
<td>Test 22108</td>
<td>0.4</td>
<td>54 51</td>
</tr>
<tr>
<td>2.1.9 Rh. typing (complete)</td>
<td>Test 22109</td>
<td>3.5</td>
<td>472 n/a</td>
</tr>
<tr>
<td>2.1.10 Direct antiglobulin test</td>
<td>Test 22110</td>
<td>0.5</td>
<td>n/a n/a</td>
</tr>
<tr>
<td>2.1.11 Direct antiglobulin test (gel test)</td>
<td>Test 22111</td>
<td>1.0</td>
<td>135 n/a</td>
</tr>
<tr>
<td>2.1.12 Cross matching</td>
<td>Test 22114</td>
<td>0.8</td>
<td>108 103</td>
</tr>
<tr>
<td>2.1.13 Cross matching (gel test)</td>
<td>Test 22115</td>
<td>1.5</td>
<td>202 n/a</td>
</tr>
</tbody>
</table>

RH = regional hospital; PH = provincial hospital; DH = district hospital, THB = Thai baht, n/a = not available

**Table 2. Data on direct nonmedical costs for outpatients per visit**

<table>
<thead>
<tr>
<th>Data</th>
<th>Mean (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HC</td>
</tr>
<tr>
<td>Distance from home to health facilities (km)</td>
<td>3.85 (0.28)</td>
</tr>
<tr>
<td>Time spent from home to health facilities (min)</td>
<td>18 (0.72)</td>
</tr>
<tr>
<td>Time spent for receiving service, including</td>
<td>69 (3.10)</td>
</tr>
<tr>
<td>traveling (from home to home) (min)</td>
<td></td>
</tr>
<tr>
<td>Traveling cost*</td>
<td>53.72 (3.53)</td>
</tr>
<tr>
<td>Meal cost*</td>
<td>13.36 (1.81)</td>
</tr>
<tr>
<td>Patient real income loss*</td>
<td>13.71 (3.67)</td>
</tr>
<tr>
<td>Accompanying person real income loss*</td>
<td>5.76 (1.96)</td>
</tr>
</tbody>
</table>

SE = standard error, HC = health center; RH = regional hospital; PH = provincial hospital; DH = district hospital, km = kilometer, min = minute

* Thai baht, 2009 values
economic evaluations more convenient, faster, and more reliable for national policy decision-making. The next revision should be developed on the back of the recommendations suggested herein.

Acknowledgement
The author wish 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. The author wishes to thank graduate students in Pharmacy Administration at the Faculty of Pharmacy, Mahidol University for participating in the project and Professor Peter Coyte, Department of Health Policy, Management and Evaluation, Faculty of Medicine University of Toronto, Canada for giving suggestions on manuscript preparation.

Potential conflicts of interest
None.

References


รายการคนทุนมาตรฐานเพื่อการประเมินความคุ้มค่าทางสาธารณสุขในประเทศไทย

อาจารย์วิศิษฐ์

โครงการนี้เริ่มต้นในพื้นที่ด้านหุ้นของสมาคมมาตรฐานของบริการทางการแพทย์ในสถาบันบริการระดับต่าง ๆ และค้นหาของผู้บริการและครอบครัวในการรับการรักษาสำหรับเด็กในภาวะที่มีความต้องการทางการดูแลสุขภาพ บริการอรักษาหุ้นต่ำในหน่วยงานของบริการทางการแพทย์โดยด้านนี้การ ...

905

3,091

2 ระดับ คือ โรงพยาบาลเอกชน และโรงพยาบาลทั่วไปหรือโรงพยาบาลสุขภาพ ซึ่งมีการมารับบริการของผู้ป่วยนอก แยกเป็นระยะทางจากบ้านของผู้บริการ เวลาที่ใช้ในการมารับบริการ ค่าดินทรายและค่าอาหาร รายการค้นหุ้นมาตรฐานที่ชัดเจนในครั้งนี้จึงจะช่วยให้การประเมินความคุ้มค่าทางสาธารณสุขสะดวกและรวดเร็วขึ้น อย่างไรก็ตามการค้นหุ้นมาตรฐานแบบจำลองนี้จะได้ถูกใช้ในการปรับปรุงครั้งต่อไป