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).

Correspondence to:
Praditsitthikorn N, Health Intervention and Technology Assessment Program, Ministry of Public Health, 6th Floor, 6th Building, Department of Health, Ministry of Public Health, Tiwanon Road, Nonthaburi 11000, Thailand.
Phone: 0-2590-4549, Fax: 0-2590-4369
E-mail: naiyana.p@hitap.net

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[3].

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[4], data of disease surveillance by the Bureau of Epidemiology[5], Cancer Registry by the National Cancer Institute[6], Burden of Disease by the International Health Policy Program[7] and Health Statistics by the Bureau of Policy and Strategy[8].

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+).

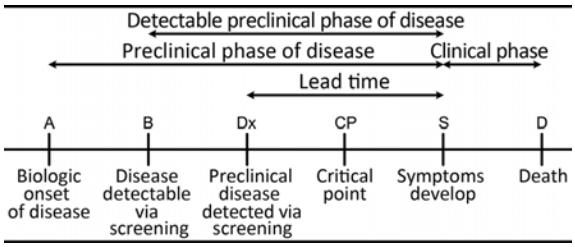
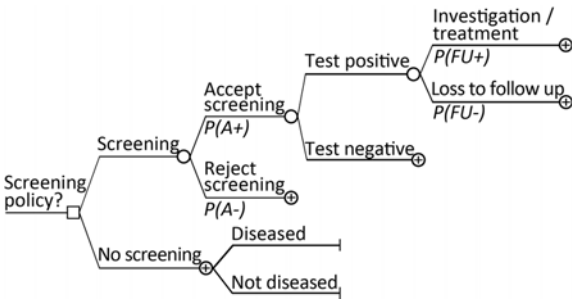


Fig. 1 Natural history of disease[3].



P(A+) = Probability of people accept the screening test, or acceptance rate.
P(A–) = Probability of people reject the screening test, or 1-acceptance rate.
P(FU+) = Probability that people complete the further in vestigation and/or treatment, or follow-up rate.
P(FU–) = Probability that people loss from follow-up, or 1-follow-up rate.
⊕ = Additional of two branches of tree; i.e. diseased and not diseased.

Fig. 2 Example of the application of acceptance rate and loss to follow-up rate in a decision tree model.

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 know 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

Screening test results	True diagnosis		Total
	Diseased	Not diseased	
Positive	a	b	a + b
Negative	c	d	c + d
Total	a + c	b + d	a + b + c + d

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 = a/(a+c)
Specificity (SP) or P(T–|D–) is = d/(b+d)
Prevalence (p) is probability of having disease; P(D+) = (a+c)/(a+b+c+d)
Positive predictive value (PPV) or P(D+|T+) is = a/a+b
written in probability notation:

$$= \frac{P(T+|D+)P(D+)}{P(T+|D+)P(D+) + P(T+|D-)P(D-)} \quad \text{or} \quad \frac{SE \times p}{(SE \times p) + [(1-SP) \times (1-p)]}$$

Negative predictive value (NPV) or P(D–|T–) is = d/c+d
written in probability notation:

$$= \frac{P(T-|D-)P(D-)}{P(T-|D-)P(D-) + P(T-|D+)P(D+)} \quad \text{or} \quad \frac{SP \times (1-p)}{[SP \times (1-p)] + [(1-SE) \times p]}$$

Probability of a positive test result or P(T+) is = (a+b)/(a+b+c+d)
written in probability notation:
= P(T+|D+)P(D+) + P(T+|D–)P(D–) or = (SE x p) + [(1–SP) x (1–p)]
Probability of a negative test result or P(T–) is = (c+d)/(a+b+c+d)
written in probability notation:
= P(T–|D+)P(D+) + P(T–|D–)P(D–) or = [(1–SE) x p] + [SP x (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.

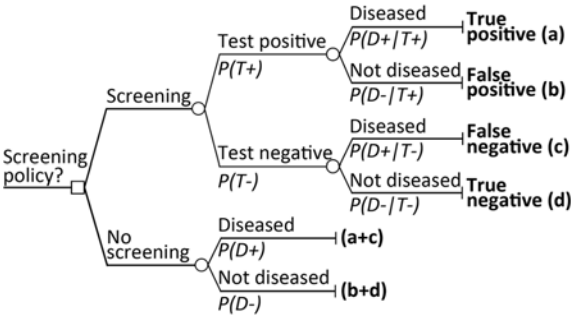
Performing tests in series is remarkably cost-efficient when screening for rare conditions and often used when the second test is expensive and/or risky. With the OR rule, if the first test is positive, the diagnosis is positive; otherwise, the second test is performed. If the second test is positive 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

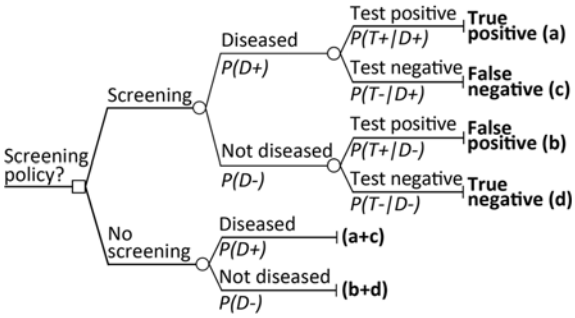
Decision rules			Accuracy of combined test
Test A	Test B	Diagnosis	
Parallel or simultaneous testing, OR rule			
Either test positive		Positive	$SE_c = SE_a + SE_b - (SE_a \times SE_b)$
Negative	Negative	Negative	$SP_c = SP_a \times SP_b$
Parallel or simultaneous testing, AND rule			
Positive	Positive	Positive	$SE_c = SE_a \times SE_b$
Either test negative		Negative	$SP_c = SP_a + SP_b - (SP_a \times SP_b)$
Series or sequential testing, OR rule			
Positive	Not performed	Positive	
Negative	Positive	Positive	$SE_c = SE_a + (1 - SE_a) \times SE_b$
Negative	Negative	Negative	$SP_c = SP_a \times SP_b$
Series or sequential testing, AND rule			
Positive	Positive	Positive	
Positive	Negative	Negative	$SE_c = SE_a \times SE_b$
Negative	Not performed	Negative	$SP_c = SP_a + (1 - SP_a) \times SP_b$

SE_a = sensitivity of test A alone; SE_b = sensitivity of test B alone; SE_c = overall sensitivity; SP_a = specificity of test A alone; SP_b = specificity of test B alone; SP_c = overall specificity



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

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



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

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

ordered (Fig. 3) or by disease status (Fig. 4).

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 Fig. 1); 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

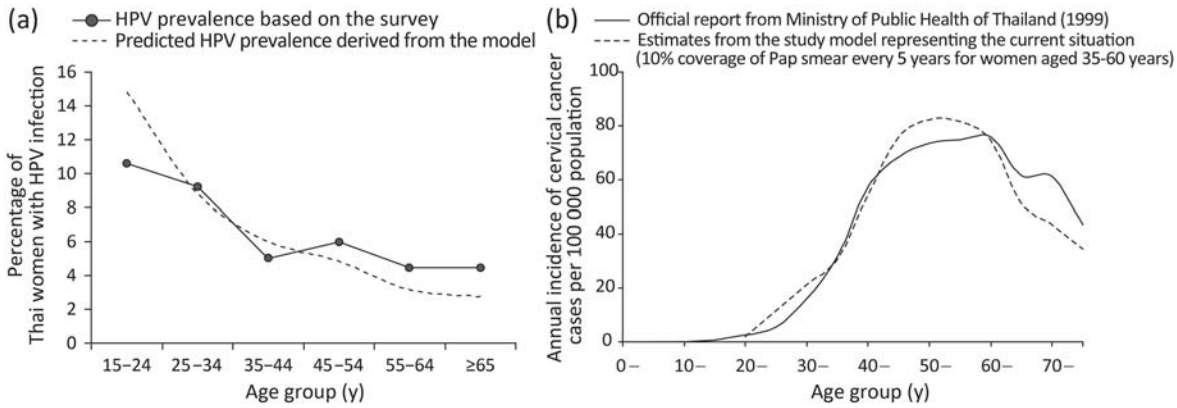


Fig. 5 Example of model validation: comparing outputs from model prediction against observed/population data[13].

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 (Praditsitthikorn 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

1. Commission on Chronic Illness. Prevention of chronic illness. Chronic illness in the United States: Volume I. Cambridge, MA: Harvard University Press; 1957: 45.
2. Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: World Health Organization; 1968.
3. Herman CR, Gill HK, Eng J, Fajardo LL. Screening for preclinical disease: test and disease characteristics. Am J Roentgenol 2002; 179: 825-31.
4. The National Statistical Office. Population and housing census in Thailand [Internet]. 2010 [cited 2013 Nov 1]. Available from: http://popcensus.nso.go.th/
5. Bureau of Epidemiology, Ministry of Public Health. Epidemiological surveillance and investigation in Thailand [Internet]. 2013 [cited 2013 Nov 1]. Available from: http://www.boe.moph.go.th/
6. The National Cancer Institute. Cancer registry [Internet]. 2013 [cited 2013 Nov 1]. Available from: http://www.nci.go.th/th/cancer_record/cancer_rec1.html
7. International Health Policy Program. Burden of disease [Internet]. 2012 [2013 Oct 15]. Available from: http://www.thaibod.net
8. Bureau of Policy and Strategy, Ministry of Public Health. Health statistics [Internet]. 2013 [cited 2013 Nov 1]. Available from: http://bps.ops.moph.go.th/index.php?mod=bps&doc=5
9. Gray AM, Clarke PM, Wolstenholme JL, Wordsworth S. Applied methods of cost-effectiveness analysis in health care. New York: Oxford University Press; 2011.
10. Joseph L, Reinhold C. Fundamentals of clinical research for radiologists. Introduction to probability theory and sampling distributions. AJR Am J Roentgenol 2003; 180: 917-23.
11. Weinstein S, Obuchowski NA, Lieber ML. Clinical evaluation of diagnostic tests. AJR Am J Roentgenol 2005; 184: 14-9.
12. Phelps CE, Mushlin AI. Focusing technology assessment using medical decision theory. Med Decis Making 1988; 8: 279-89.
13. Praditsitthikorn N, Teerawattananon Y, Tantivess S, Limwattananon S, Riewpaiboon A, Chichareon S, et al. Economic evaluation of policy options for prevention and control of cervical cancer in Thailand. Pharmacoeconomics 2011; 29: 781-806.

การประเมินความคุ้มค่าทางสาธารณสุขของมาตรการคัดกรองโรค

นัยนา ประดิษฐ์สิทธิกร

ปัจจุบันหลายหน่วยงานหันมาให้ความสนใจกับการสร้างเสริมสุขภาพและป้องกันโรคมากยิ่งขึ้น รวมทั้งสนับสนุนให้มีการคัดกรองในประชากรอย่างเหมาะสม อย่างไรก็ตามความสำเร็จของโครงการคัดกรองขึ้นอยู่กับหลายปัจจัยทั้ง ความรุนแรงของโรค เทคนิคการคัดกรอง การบริหารจัดการโครงการ เป็นต้น ดังนั้นเพื่อให้ผู้มีส่วนเกี่ยวข้องวางแผนการให้คัดกรองอย่างสมเหตุสมผลและได้ประโยชน์สูงสุด การประเมินความคุ้มค่าด้านสุขภาพด้วยการสร้างแบบจำลองโรคย่อมมีส่วนช่วยอย่างมาก ด้วยการสร้างสถานการณ์เลียนแบบความสามารถในทางปฏิบัติและเปรียบเทียบผลกระทบที่อาจเกิดขึ้นทั้งทางสุขภาพและเศรษฐศาสตร์ เพื่อประกอบการตัดสินใจก่อนที่จะนำไปปฏิบัติจริง เนื้อหาในบทนี้จึงมีวัตถุประสงค์เพื่อให้ข้อมูลที่จำเป็นต้องทราบก่อนทำการศึกษาความคุ้มค่าของมาตรการคัดกรองใด ๆ