Health technology assessment: an introduction with emphasis on assessment of diagnostic tests

Hindrik Vondeling
Department of Health Economics
University of Southern Denmark
Odense, Denmark

Outline

- Health technology assessment
  - Background and theoretical framework
- Assessment of diagnostic tests
  - Categorization of diagnostic research
    - Diagnostic accuracy
      - Phase I, II, III and IV studies
  - Assessment of value for money of diagnostic and treatment strategies
    - Economic evaluation carried out alongside randomised controlled clinical trials

Health technology

- Definition
  - The drugs, devices and surgical procedures used in health care, and the organisational and supportive systems within which care is provided
- Examples of organisational and supportive systems
  - Primary care, secondary care, integrated care, intensive care
  - The electronic patient record, telemedicine, the ‘hotel’ function in hospitals

Health technology

- Associated with health benefits
  - Increase of life-expectancy
  - Increase in healthy life-year expectancy
  - Reduction of human suffering
- Associated with controversies
  - Technologies that do more harm than good
    - Issues of safety and efficacy
  - High cost of technology
    - Issues of ‘value for money’

Health technology assessment

- Definition: HTA aims at the systematic evaluation of the properties and effects of health technology
- The primary purpose of HTA is to provide objective, reliable, and valid information to support decisions in policy and practice at the local, regional, national and international level
- HTA uses methods from a variety of disciplines
  - E.g. study designs from clinical epidemiology (for example randomised controlled trials) and analytical methods from health economics (for example cost-benefit analysis)

Health technology assessment

- Scope
  - Aspects
    - Safety, efficacy/effectiveness, accuracy, economic, legal, organizational, ethical, educational, cultural, accreditation and certification
  - Diffusion in health care systems
    - Adoption
    - Use

HTA and decision making

- New technology
  - Industry - marketing and promotion
  - Policy makers - regulation, legislation, evaluation
  - Hospitals and providers - purchase and use
  - Patients - acceptability
  - Insurers - payment

A process or system of HTA

- Identification of technologies
- Selection of technologies most in need of assessment (priority setting)
- Assessment (primary data collection)
- Synthesis
  - Combining primary and secondary data
  - Making recommendations about appropriate use
- Dissemination of information
- Implementation in policy and practice

Overview of HTA

- HTA well-established in the Nordic Countries
  - Danish Center for Evaluation and Health Technology Assessment
  - http://www.dacehta.dk (DACEHTA)
- The Swedish Council on Technology Assessment in Health Care
  - http://www.sbu.se (SBU)
- The Norwegian Centre for the Health Services
  - http://www.nokc.no (NOKC)
- The Finnish Office for Health Technology Assessment
  - http://www.stakes.fi/finohta/ (FinOHTA)

Overview of HTA

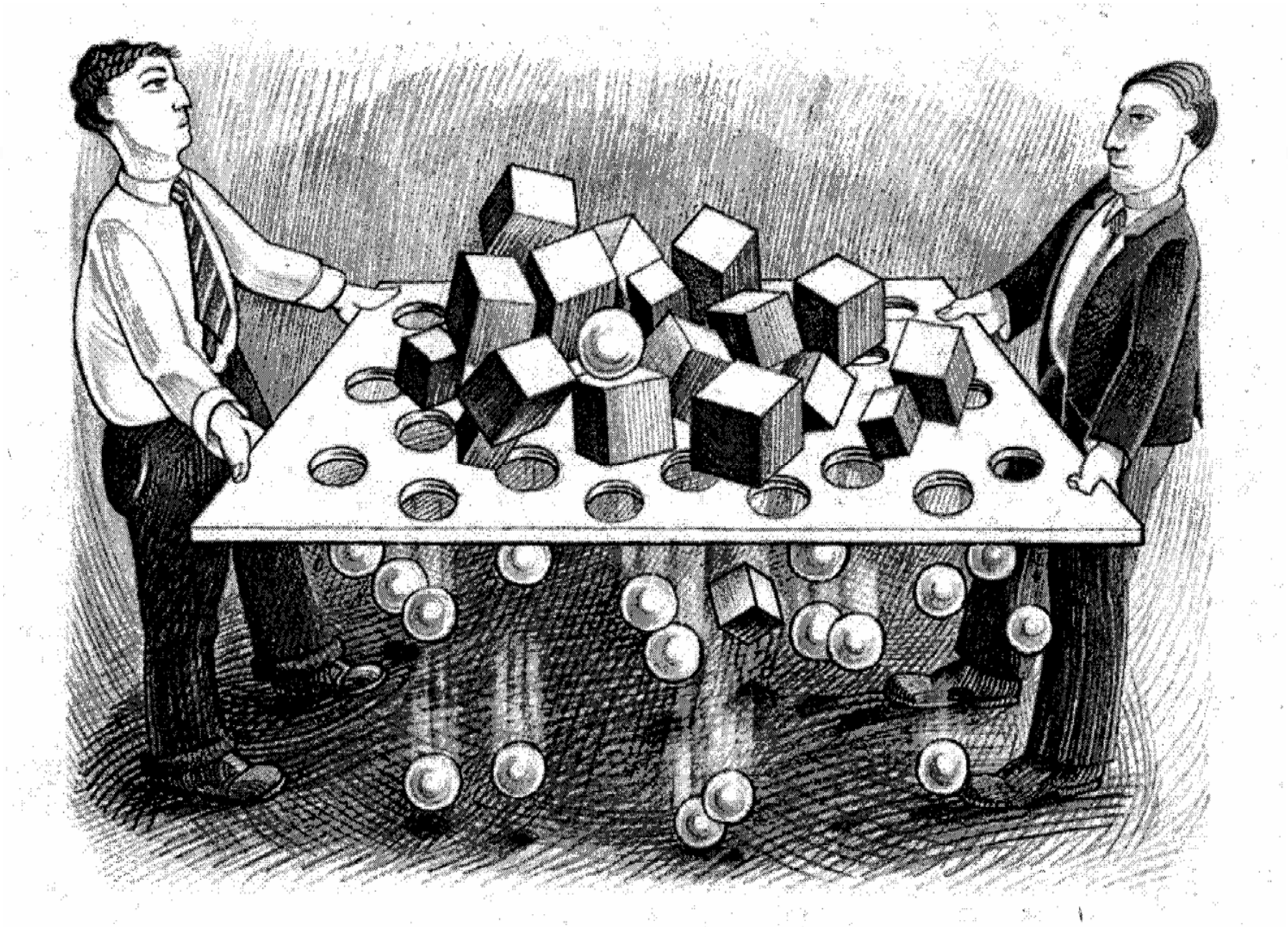
- International Association of Health Technology Assessment Agencies (INAHTA)
  - 41 member agencies in 21 countries
  - http://www.inahta.org/inahta_web/index.asp
- Health Technology Assessment International (HTA*i*)
  - http://www.htai.org

Assessing diagnostic tests

Introduction

Diagnostic technology

- Major contribution to reduction of clinical uncertainty
  - **Physicians' views of the relative importance of 30 medical innovations in the past 30 years**
    - Technologies and ranking
      - **1 MRI and CT**
      - 2 ACE inhibitors
      - 3 Balloon angioplasty
      - 4 Statins
      - **5 Mammography** (Fuchs and Sox, 2001)
- Major concern: overuse
      - Defensive medicine and patient pressures



General terminology

- Test – any method for obtaining additional information on a patient's health status
  - Patient history
  - Clinical examination
  - Laboratory tests
  - Imaging tests
  - Function tests
  - Histopathology

Diagnostic tests versus screening tests

- ’Tests performed in persons with a symptom or sign of an illness are usually termed *diagnostic*, whereas those done in individuals with no such symptom or sign are referred to as *screening*. The underlying rationale as to when a test ought to be applied, however, is identical for these two types’
  - A positive test result might lead to the induction of therapy when it might not otherwise have been considered
  - A negative test might can lead to the decision *not* to initiate therapy when it otherwise would have been given

(Weiss in Rothmann and Greenland 1998)

Categorization of diagnostic research in relation to study objectives

- *1 Assessing diagnostic accuracy*
  - Increasing certainty on the presence or absence of disease
- *2 Assessing the impact of (additional or replacing) diagnostic testing on clinical management*
  - E.g. determining presence, localisation and shape of arterial lesions is necessary for treatment decisions
- 3 Assessing the impact of (additional or replacing) diagnostic testing on prognosis
  - E.g. as a starting point for clinical followup and informing patients

(adapted from Knottnerus et al 2002)

Categorization of diagnostic research in relation to study objectives

- *4 Determining the most (cost)-effective diagnostic (and treatment) strategy*
  - Scarcity of resources
- 5 Monitoring clinical course
  - When a disease is untreated, or during or after treatment
- 6 Measuring fitness
  - For example, for sporting activity or for employment
- 7 Assessing (synthesising) results of multiple studies
- 8 Translating findings into practice and policy

(adapted from Knottnerus et al 2002)

Assessing diagnostic accuracy

Introduction

Terminology in studies of diagnostic accuracy

- Index test: The test under evaluation
- Reference standard: The best available method for establishing the presence or the absence of the condition of interest
  - A single method
  - A combination of methods
    - Laboratory tests, imaging tests and pathology, dedicated clinical followup of participants
- Accuracy: The amount of agreement between the index test and the reference standard

Four types of diagnostic research questions that should be consecutively positively answered

- Phase I - Do patients with the target disorder have different test results from normal individuals?
- Phase II - Are patients with certain test results more likely to have the target disorder than patients with other test results?
- Phase III - Among patients in whom it is clinically sensible to suspect the disorder, does the level of the test result distinguish those with and without the disorder?
- Phase IV - Do patients who undergo the test fare better (including ultimate health outcomes) than similar patients who do not?

(Sackett and Haynes, 2002)

Cross sectional design of diagnostic accuracy studies

- Condition: the results of the index test and reference standard should be known for all subjects in the study population
- Case control sampling or case-referent design (phase I)
  - Comparing test distributions in samples already known to have the disorder (cases) and known to be free of it (controls)
- Sampling based on test results (phase I)
  - Comparing disease distributions in samples with already known test results
- Surveys in ’indicated’ population
  - Survey in a target population in which testing would be relevant

(Knottnerus et al 2002)

STARD initiative

- Standards for Reporting of Diagnostic Accuracy
- 25 recommendations
- Aimed at promoting complete and accurate reporting of studies of diagnostic accuracy
  - Allows to detect the potential for bias in the study
    - Internal validity
  - Allows to assess the generalizability and applicability of the results
    - External validity

Recommendation 21 on reporting of estimates

- Report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)

Measures of diagnostic accuracy

- *Sensitivity*
- *Specificity*
- *Positive and negative predictive value*
- The area under a receiver operator characteristic (ROC) curve
- Likelihood ratios of postive and negative test results
- Diagnostic odds ratio’s

Measures of diagnostic test performance (I)

		Reference test	
		+	-
Index test	+	a	b
	-	c	d

a true positive - correct positive test result

b false positive - incorrect positive test result

c false negative - incorrect negative test result

d true negative - correct negative test result

Measures of diagnostic test performance: sensitivity

		Reference test	
		+	-
Index test	+	a	b
	-	c	d

a / a + c sensitivity (proportion of people with the target disorder who have a positive test result)

How good is this test at picking up people with the condition?

Measures of diagnostic test performance: specificity

		Reference test	
		+	-
Index test	+	a	b
	-	c	d

d / b + d specificity (proportion of people without the target disorder who have a negative test result)

How good is this test at correctly excluding people without the condition?

Sensitivity and specificity of tests in clinical practice

Test	Sensitivity (%)	Specificity (%)
Coronary stenosis		
Exercise electrocardiography	65	89
Stress thallium scintigraphy	85	85
Pancreatic cancer		
Ultrasonography	70	85
Computed tomography	85	90
Angiography	75	80

Measures of diagnostic test performance: positive predictive value

		Reference test	
		+	-
Index test	+	a	b
	-	c	d

a / a + b positive predictive value (PPV; the probability of disease among all persons with a positive test result)

If a person tests positive, what is the probability that he or she has the condition?

Measures of diagnostic test performance: negative predictive value

		Reference test	
		+	-
Index test	+	a	b
	-	c	d

d / c + d negative predictive value (NPV; the probability of non-disease among all persons with a negative test result)

If a person tests negative, what is the probability that he or she does not have the condition?

Example

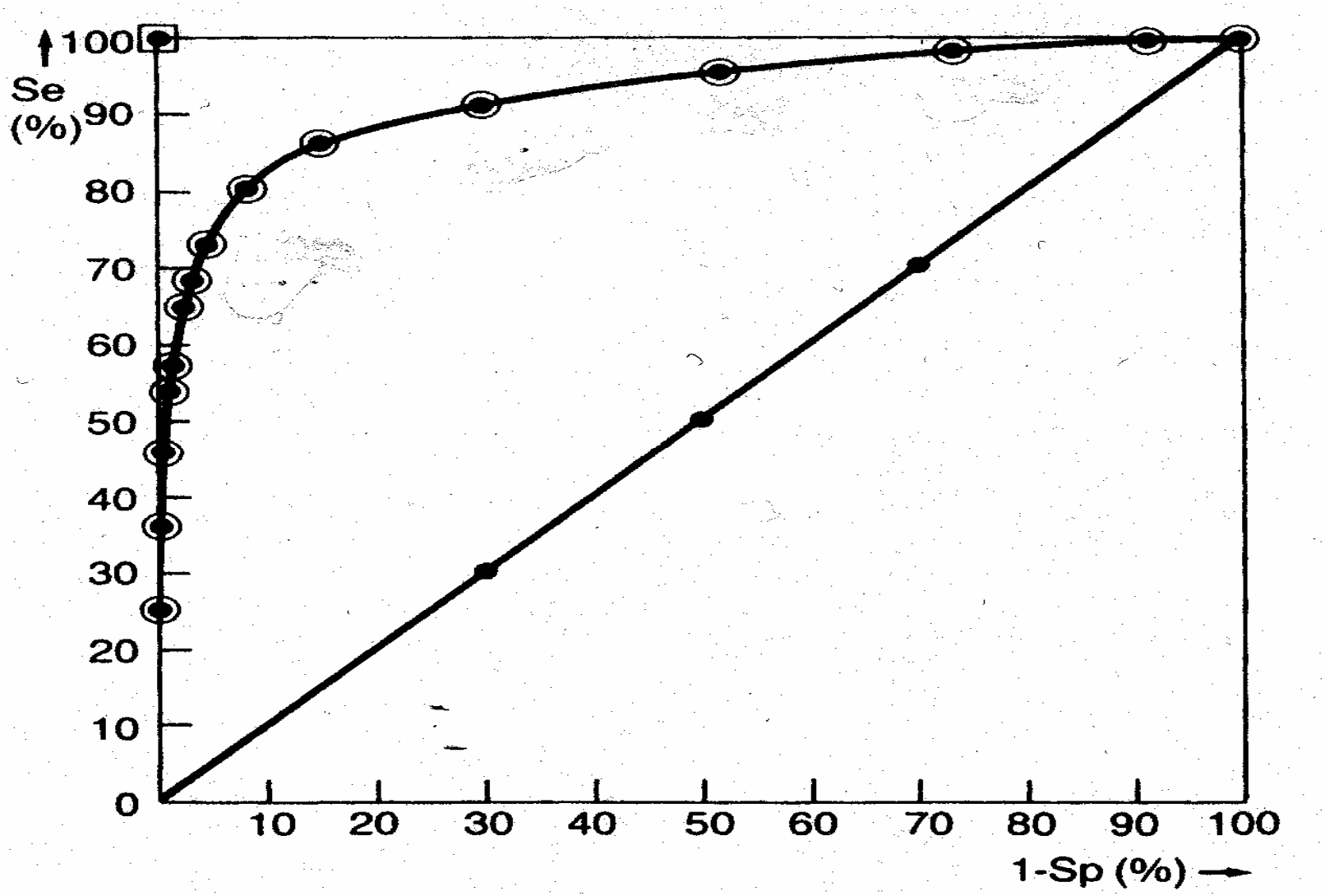
		Reference test	
		+	-
Index test	+	a = 950	b = 200
	-	c = 50	d = 800

Feature	Formula	Data	Value (%, 95% CI)
Sensitivity	a / a+c	950 / 1000	95 (94 - 96)
Specificity	d / b+d	800 / 1000	80 (77 - 82)
PPV	a / a+b	950 / 1150	83 (80 – 85)
NPV	d / c+d	800 / 850	94 (92 – 96)

Measures of diagnostic test performance (VI)
Receiver operating characteristic (ROC) curve

- A ROC curve represents the relationship between the 'true-positive' fraction (sensitivity) and the false-positive' fraction (1-sensitivity). It displays the trade-offs between sensitivity and specificity as a result of varying the cut-off value for positivity in case of a continuous test result
- The total area under the curve is 1
- A test with an area under the curve of 0.5 does not provide diagnostic evidence

ROC Curve



Measures of diagnostic test performance (VII) Likelihood ratio's

- Likelihood ratio of a positive test
  - sensitivity / (1 – specificity)
  - How much more likely is a positive test to be found in a person with the condition than in a person without it?
    - Value: usually >1

Measurement of diagnostic test performance (VIII) Likelihood ratio's

- Likelihood ratio of a negative test
  - (1 – sensitivity) / specificity
  - How much more likely is a negative test to be found in a person without the condition than in a person with it?
    - Value: usually between 0 and 1

Measurement of diagnostic test performance (IX)
Diagnostic Odds Ratio (DOR)

		Reference test	
		+	-
Index test	+	a	b
	-	c	d

Overall (single indicator) measure of diagnostic accuracy

DOR: ad / bc

The odds of positivity of diseased persons, divided by the odds of positivity among non-diseased.

Value: if DOR = 1 the test does not provide diagnostic evidence

Assessing the appropriateness of testing and its impact on clinical management

Design issues and examples

Design of studies assessing the impact of diagnostic tests on patient management

- Randomised controlled clinical trials
- Observational studies
  - Cohort study
  - Case-control study
  - Before-after study

Acceptability of tests (I)

’If the probability of disease is extremely low or high, the outcome of subsequent investigations rarely influences management and false positive or false negative results, respectively, are common’

’… tests with moderate specificity are inappropriate for population screening (with low probability of disease) because of the high risk of false positive results’

(Knottnerus et al 2002)

Acceptability of tests (II)

’The Guthrie heel-prick screening test for congenital hypothyroidism, peformed on all babies in Britain soon after birth, is over 99% sensitive but has a positive predictive value of only 6% (it picks up almost all babies with the condition at the expense of a high false positive rate), and rightly so. It is more important to pick up every baby with this treatable condition who would otherwise develop severe mental handicap than to save hundreds the minor stress of a repeat blood test’

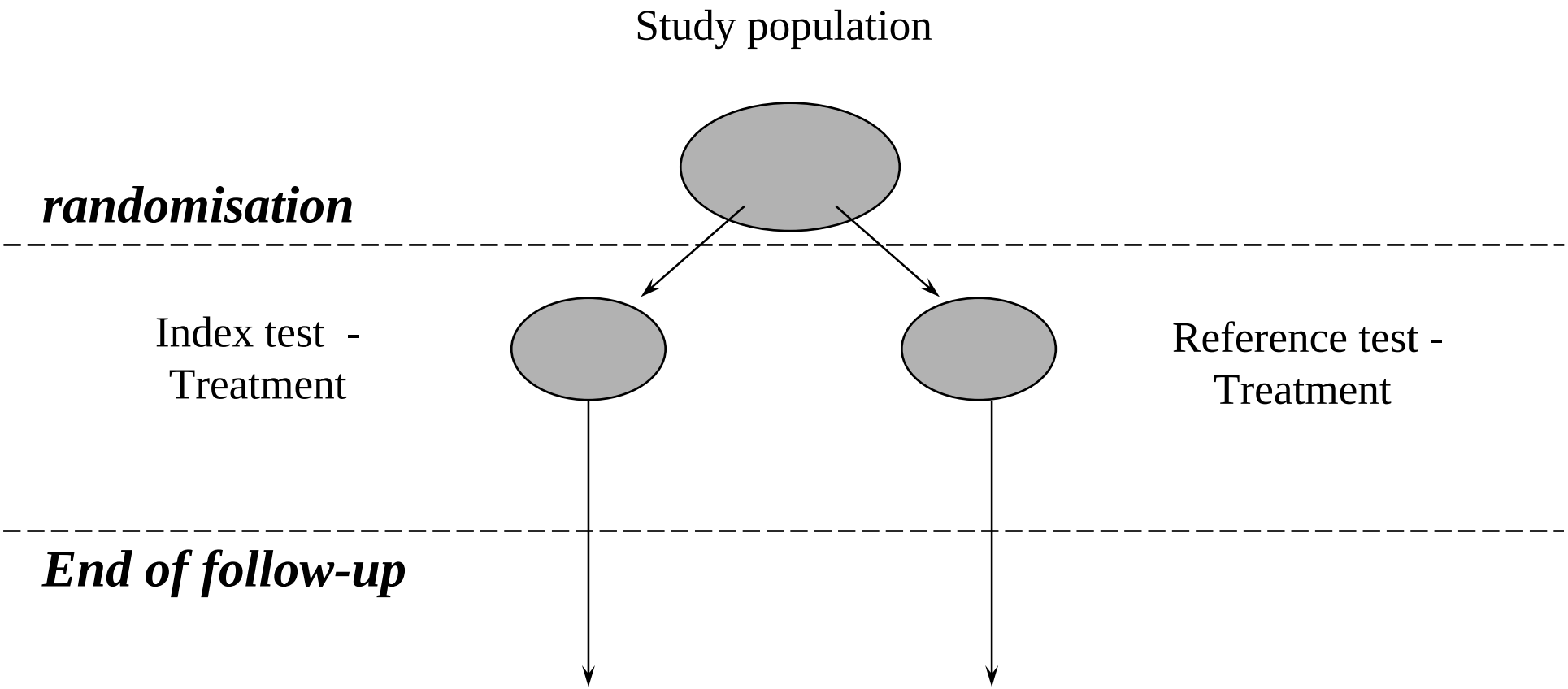
(Greenhalgh 1997)

Assessing value for money of diagnostic and treatment strategies

Incorporating economic evaluation studies in randomised controlled clinical trials

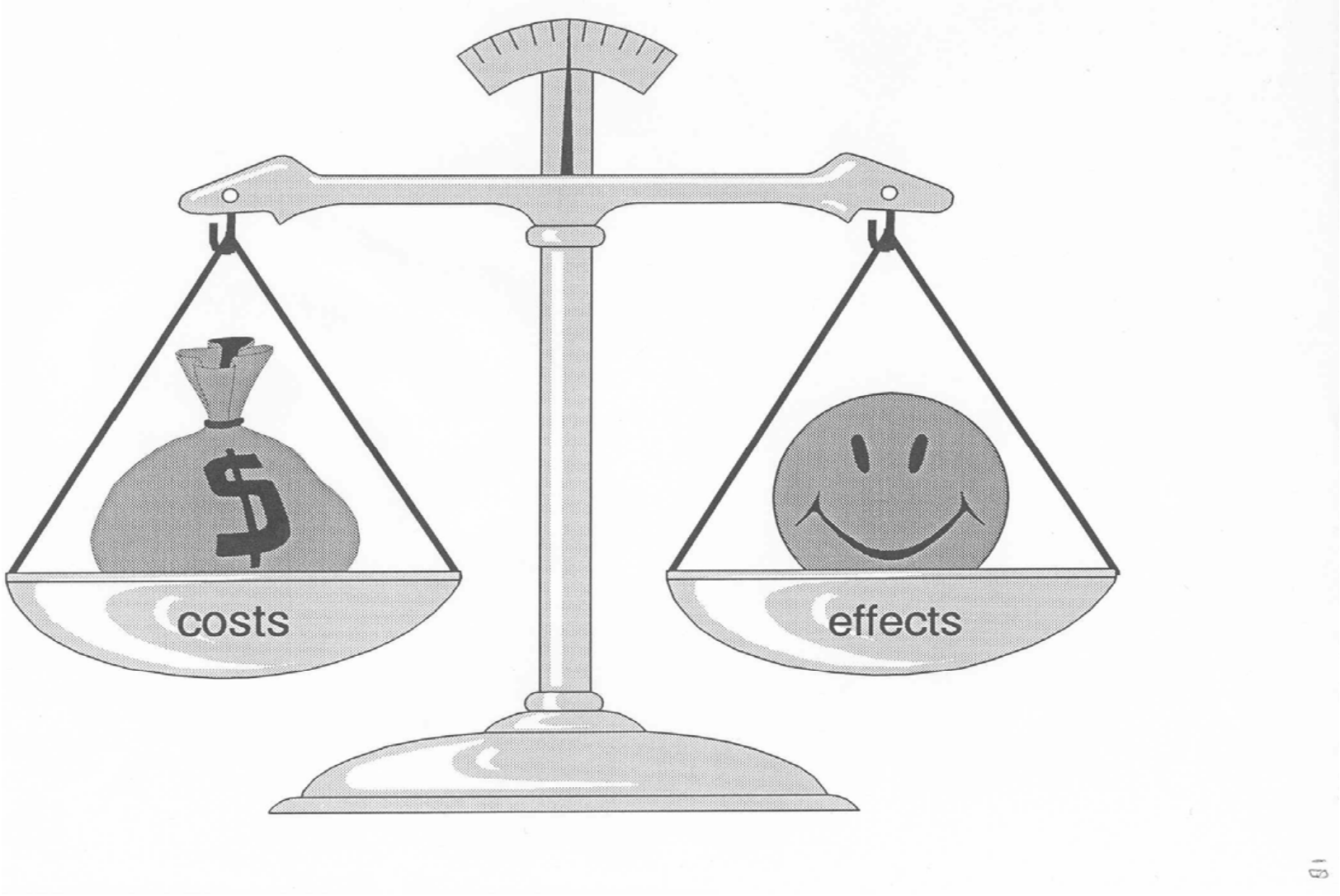
Example

Assessment of effects and costs of diagnosis and treatment in randomised clincial trials



Outcomes: Mortality, morbidity, quality-of-life
Costs: health care; patients and family; in other sectors

Economic evaluation of health care programmes



Full economic evaluations

- Are both costs (inputs) and consequences (outputs) of the alternatives examined? Yes
- Is there a comparison of two or more alternatives? Yes
  - Cost-minimization analysis (CMA)
  - Cost-effectiveness analysis (CEA)
  - Cost-utility analysis (CUA)
  - Cost-benefit analysis (CBA)

Cost-minimization analsyis

- If a common outcome of interest of two programmes is achieved at the same degree, the economic evaluation is essentially a search for the least cost-alternative

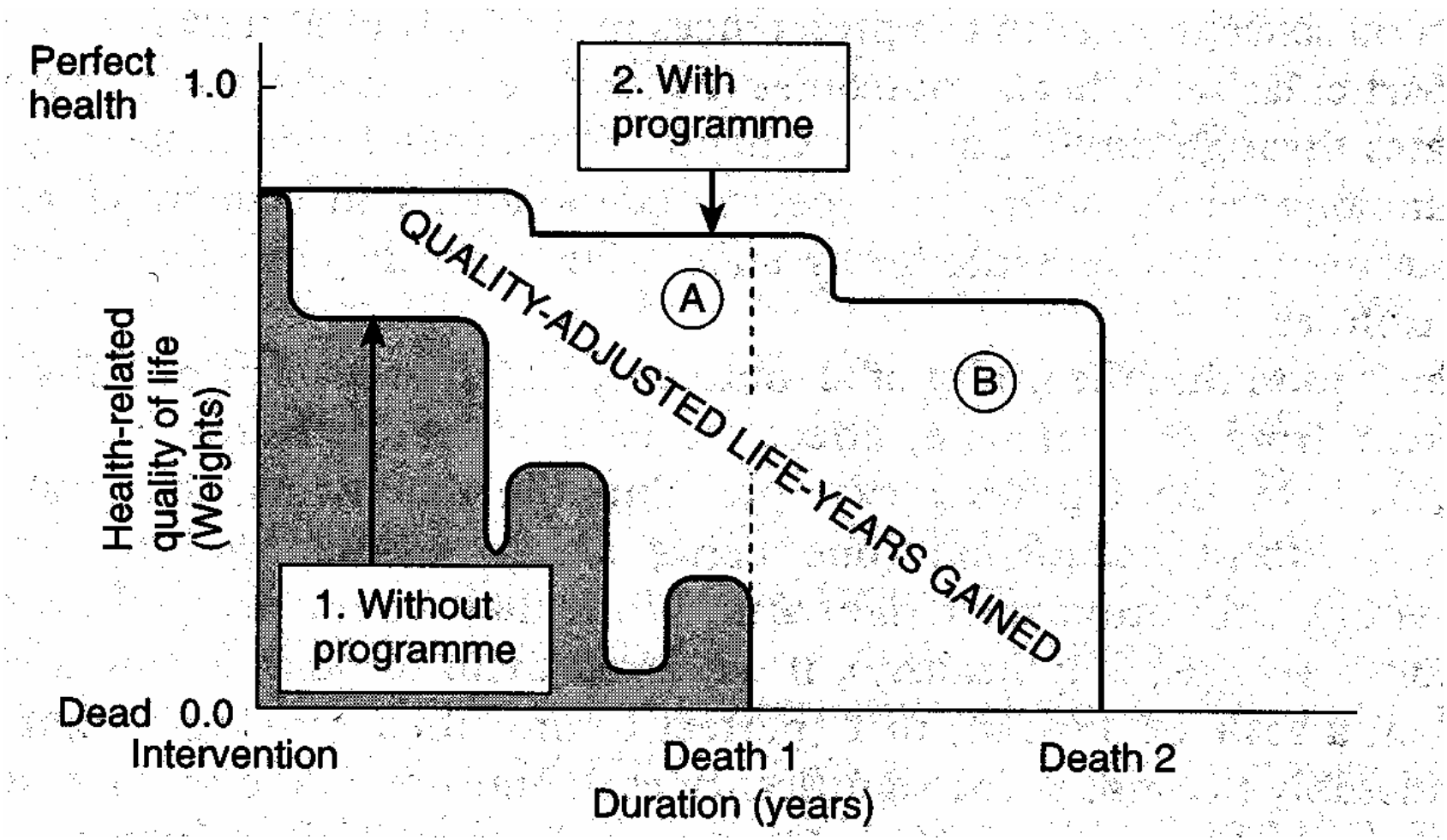
Cost-effectiveness analysis

- Analyses in which costs are related to a single, common effect which may differ in magnitude between the alternative programmes
- Cost per case detected
- Cost per life-year gained

Cost-utility analysis

- Utilities
  - Preferences that individuals or society may have for any particular set of health outcomes
  - Are usually expressed on a 0-1 scale, with 0 representing ’dead’ ad 1 representing ’perfect health’
- QALY’s
  - Adjusting the length of time affected through the health outcome by the utility value of the resulting level of health status
    - E.g. 5 year spent in a health state with a utility of 0.8 results in 4 QALY’s
- Results expressed as Costs per QALY

Cost-utility analysis



Drummond et al., 1997

Cost-benefit analysis

- Analyses which measure both the costs and consequences of alternatives in dollars (or any other monetary units)
- Method: Assessing individuals' willingness to pay (WTP) for health benefits
- Expression: a sum representing the net benefit or net loss of one programme over another

The most appropriate form of analysis is dependent on

- The problem
- The practical measurement challenges
  - The estimation methods used
- The decision to be supported
  - Allocation of resources within and/or outside the health care sector
    - Within and outside: CBA
    - Within, broad choices: CUA
    - Within, limited choices: CEA
    - Within, very limited choices: CMA

Example

- Suppose the incremental (additional) Cost per QALY of a patient management strategy including diagnosis and treatment is 20.000 US$
- Suppose societies' maximum willingness to pay for a QALY is 50.000 US$
- Decision: adopt the technology

Summary points

- Health technology assessment is a multidisciplinary and internationally oriented activity
- Diagnostic tests versus screening tests
- Diagnostic accuracy studies versus (among others) studies evaluating diagnosis as part of a patient management strategy in- or excluding issues of ‘value for money’
- Diagnostic technology: rapid technological change, accompanied by rapid changes in assessment methodology