Guide

How to judge a proposal for a screening programme

May 2004

For a list of ANAES' publications, please write to the address below with your details, or go to our website: www.anaes.fr

This document was produced in May 2004. It may be ordered (including carriage) from:
ANAES (Agence nationale d'accréditation et d'évaluation en santé)
Service communication
2, avenue du Stade de France – 93218 Saint-Denis La Plaine CEDEX – Tel.: +331 55 93 70 00 – Fax: +331 55 93 74 00

ISBN: Price:

Synopsis

Title	How to judge a proposal for a screening programme
Publication date	May 2004
Type of document	Methodological guide
Intended for	ANAES project managers and other individuals or bodies wishing to conduct an appraisal of a proposed screening programme
Requested by	ANAES
Produced by	ANAES – French National Agency for Accreditation and Evaluation in Healthcare (Health Technology Assessment and Economic Evaluation Departments)
Objectives	To produce a standard method for appraising proposals for screening programmes.
Assessment method	Critical review of the literature and discussions within two working groups
Literature search	Period: indefinite (see Annex 3)
Project management	Project leader: Dr. Emmanuel Corbillon (Department head: Dr. Bertrand Xerri); Anne-Isabelle Poullié (Department head: Catherine Rumeau-Pichon) Literature search: Emmanuelle Blondet (Department head: Rabia Bazi) Secretarial work: Sabrina Missour
Participants and validation (see Annex 1)	Working groups (in house and external) Peer reviewers Validation by ANAES Scientific Council on May 2004
ANAES appraisals of screening programmes	- Assessment of human papilloma virus (HPV) testing in primary screening for cervical cancer in France (May 2004) - Clinical and economic assessment of screening for HFE1 haemochromatosis in 2004 (April 2004) - Virtual colonoscopy in screening for colorectal cancer (Jan. 2001) - Laboratory methods for screening and diagnosis of dyslipidaemia in primary prevention (Jan. 2000)

Contents

Summary

Aim

The aim of the study was to establish guidelines for assessing a screening programme proposal.

Methods

We used an established health technology assessment method combining a critical appraisal of the literature and the opinion of a panel of experts to draw up the guidelines. We searched the Medline database (*National Library of Medicine*, United States) over an unlimited period, consulted relevant websites, and sought out the relevant grey literature. A draft report was written and submitted first to in house project managers, then to a working group of 13 members recruited from the learned societies concerned. Once amended, the report was submitted for review by 8 peer reviewers from a variety of backgrounds.

Results

The study provided answers to three main questions:

1. *How to programme the appraisal of a screening programme?*
   Three points have to be addressed before performing an appraisal. It is necessary to (i) define the objectives of the appraisal to ensure its relevance, (ii) search the literature on the topic and (iii) consult experts in the field. It may also be useful to interview patient associations and the companies that market detection kits.
2. *What data should be collected and what criteria should be used for the appraisal?*
   The relevance of a screening programme is judged by a list of criteria related to the nature of the disease to be screened for, the properties of the screening test, the diagnosis and the treatment of the disease. It is also judged by criteria related to the efficacy and safety, economic implications, administrative arrangements and assessment of the screening programme.
3. *When to perform economic modelling studies?*
   Methodological or practical problems may preclude studies "in the field". Mathematical simulation can then be used to make a synthesis of information from different sources, represent interactions among data and measure the effects of the intervention under study (clinical decision analysis).

Conclusions

The above procedure is founded on a structured and systematic approach. It encourages the taking of decisions on the basis of objective rather than subjective criteria and reduces the uncertainties in the decision-making process. The proposed assessment methods are, in principle, rigorous and well suited to solving current problems. However, these guidelines will probably undergo some changes as follow-up results become available for ongoing screening programmes, in France and elsewhere. The guidelines are also likely to evolve in response to the specific issues related to genetic screening. Their acceptance by public health decision-makers and the medical profession is subject to caution and, for example, influenced by the way the principle of precaution is viewed by decision-makers and the public.

I. Introduction

The World Health Organisation (WHO) describes screening *as the systematic application of a standard test or tests to detect a potential disease in a person who has no known signs or symptoms of that disease at the time the test is performed.* Screening tests should distinguish individuals who are apparently in good health but who probably have a given disease, from those who probably do not. The subpopulation of individuals with a higher probability of having the disease, once identified, will undergo diagnostic investigations followed by intervention. The term “intervention” covers treatment, preventive measures or information considered to be important for the affected individual.

I.1. Why this guide was produced

An appraisal of a proposed public health initiative is a detailed analysis of all the issues governing the implementation and assessment of that initiative. Because a screening programme is aimed at individuals who are symptom-free or in apparent good health, programme implementation is always preceded by an appraisal to establish whether benefits outweigh any disadvantages. Decision-makers need this information when deciding whether to implement a programme. The information must also be available to the general public to help them decide whether to participate in the programme.

An appraisal may be followed by an ongoing or intermediate assessment and then by a final study (1) (Fig. 1).

One of the tasks of the French National Agency for Accreditation and Evaluation in Healthcare (ANAES) is to identify state-of-the-art preventive, diagnostic and treatment strategies in medicine. Since 1995, ANAES and its predecessor ANDEM (the French National Medical Evaluation Agency) have produced 17 reports assessing the appropriateness of introducing a screening programme or describing the best way of organising such a programme. The methods used in these reports varied considerably. In order to move towards a standard method to meet the rising number of requests, we decided to produce a guide on how to appraise proposals for screening programmes.

I.2. How the guide was produced

The guide is based on a review of the international literature and on discussions within in house and external working groups.

I.3. Who will use the guide and how

The guide was designed to assist ANAES project managers in producing appraisals of screening programmes but may also be used as a basis for discussion and as an information source by other individuals or bodies wishing to conduct an appraisal. It consists of:

- a reminder of basic definitions and concepts;
- four sections dealing with the stages of an appraisal, from preparatory work to producing the final report (Fig. 2).

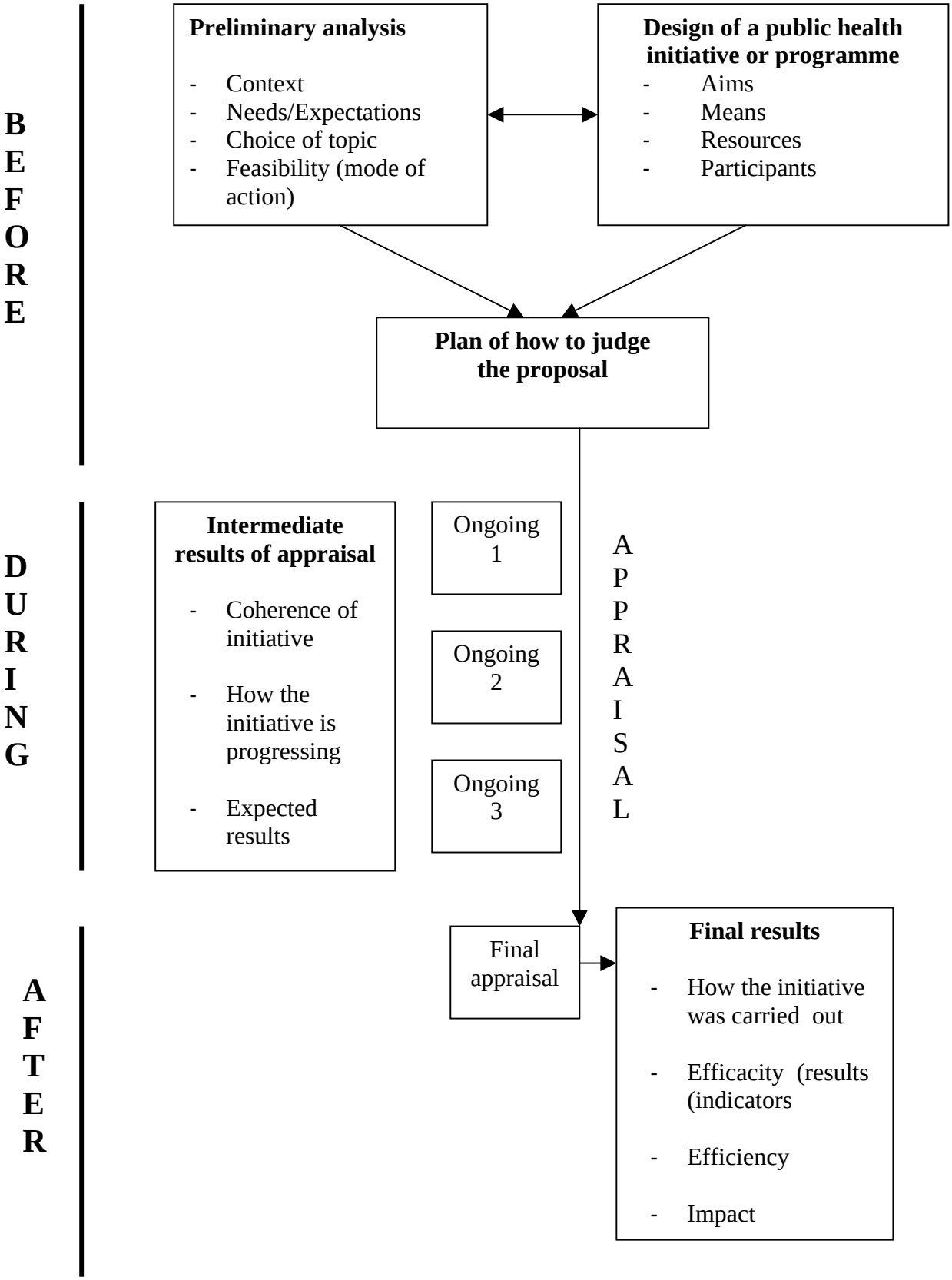

Fig. 1. Scheme for an appraisal of a public health initiative (1)

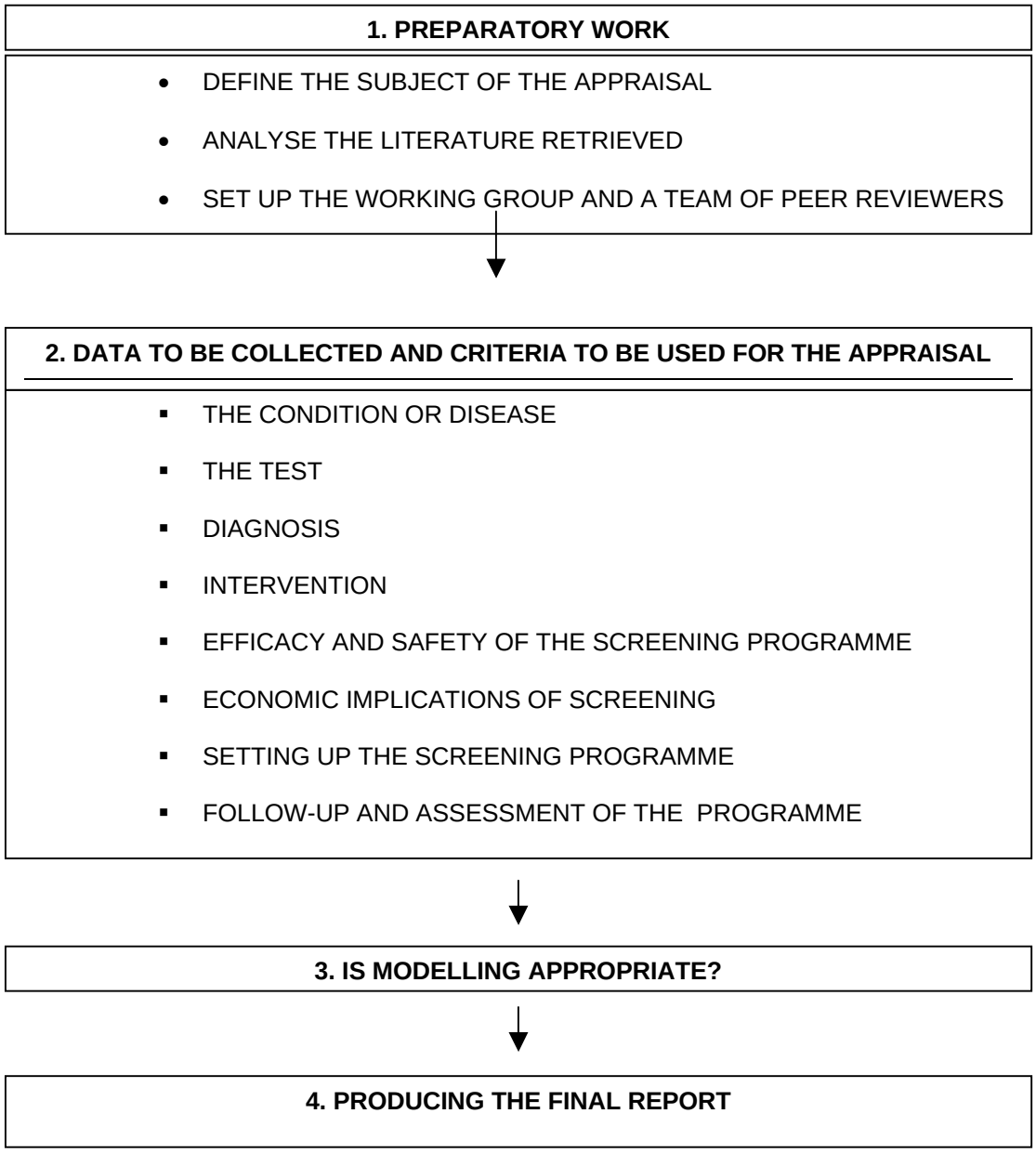

Fig. 2. The four stages of an appraisal outlined in this guide

I.4. Reminder of definitions and concepts

• Types of screening

Published articles do not always provide the same definitions for different types of screening. The literature was searched for definitions. The most recent official source was the publication of the WHO criteria in 1970 (see Table 1).

Table 1. Types of screening

Type of screening	Definition
Systematic or **mass**	The population recruited is not selected. If there is an age criterion, screening is regarded as applying to all individuals in the relevant age range
Selective or **targeted**	The population recruited is selected on previously defined criteria (risk factors revealed by controlled trials)
Community	The population is recruited from within the community. Screening is carried out as part of a specific campaign and is based on voluntary participation
Opportunistic	People are recruited for screening when they use healthcare resources (i.e. a hospital, doctor's surgery, health centre or screening centre, or when they see a company doctor)
Multiple	A battery of tests is used to screen for a number of diseases or conditions

• Screening is not diagnosis

Screening is not carried out in response to an explicit request from members of the public. A health professional takes the initiative for carrying out a clinical or laboratory test on individuals who are presumed to be in good health because they have no signs of the disease. Screening is the step before definitive diagnosis (Table 2).

Screening tools are not always the same as diagnostic tools. They have to be able to be used without risk in large populations, at a low unit cost. They are used in a general population to select individuals who have a specific condition and to differentiate the probably healthy from the probably ill, albeit with some margin of error. The probably ill will be referred to doctors and will undergo further tests before a diagnosis is either established or eliminated.

Table 2. Main differences between screening and diagnostic tests
After Durand-Zaleski, 2000 (2).

Screening test	Diagnostic test
The step before definitive diagnosis	Must provide a definite diagnosis (specific tests)
Carried out in individuals who do not appear to have the disease	Applied to individuals with defined disorders
Carried out in groups of individuals at high risk	Basically a test for an individual
Not an aid to deciding treatment	May be used as a second-line test after screening
	Leads to a treatment decision

I.5. When to appraise a screening programme

A screening programme is a public health initiative, i.e. a list of steps to be taken to improve the health of a community (1). Because the implementation of widespread screening programmes raises major ethical, legal, social, medical, organisational and economic problems which require initial and ongoing evaluation (3), the decision-maker should assess whether introducing a screening programme is appropriate and whether its aims respond to needs, i.e. a screening programme should be assessed when designed.

Screening benefits individuals who are correctly identified as positive or negative.

- *For true positives*, the advantages may be a more effective intervention (treatment or preventive measure) started earlier, a better chance of survival, or an improvement in quality of life from less treatment. There may be savings in resources due to a lower total cost of managing the disease (less radical initial treatment and reduced morbidity and mortality).
- *For true negatives*, the advantages may be patients feeling reassured and possibly reduced monitoring.

In addition, screening can contribute to equity by providing access to care for individuals who would not have benefited from this care if they had not been involved in the programme.

II. Preparing an appraisal of a screening programme

Three steps precede the collection of the information needed to appraise a screening programme.

Step 1. Define the subject of the appraisal

The first step is to identify aims and stakeholders, and to understand the issues involved. The appraisal plan should follow a logical path. It should:

- begin with the issues underlying the appraisal
- produce a set of questions (Box 1)
- state how answers will be found.

This is the pre-appraisal goal-setting stage ("target setting meeting") that establishes the relevance of the appraisal.

Box 1. Questions to be answered before the start of an appraisal (4-6)
Some items were raised by the working group and were not based on the articles cited.

(i) Who has requested the appraisal and why?

The organisation requesting an appraisal is usually a government body or authority with responsibility for health policy, but it may also be:

- a patients' or users' association
- a learned medical society, university, or expert
- a medical professional union.

Identifying stakeholders (the body making the request, the health professionals involved), reviewing the steps motivating an appraisal and the reasons given are key factors in understanding the aims of an appraisal, anticipating problems and determining priorities.

(ii) Is there a background file describing the institutions, partners, practices, technologies involved?
The official aims are sometimes stated in health policy papers. These should be compared with any document that could shed light on the decision-makers' intentions, such as statements by political leaders or administrative executives.
A short, simple synopsis should describe the background to the issue, any ongoing or published studies and the current situation regarding individual screening.

(iii) Who are the appraisal results intended for (other than those making the request)?
The type of results and practical consequences, along with the potential impact on the target population, need to be identified in the light of the context and expectations of the request. Targets are, for instance, healthcare administration professionals, doctors in non-hospital practice, university medical departments, the general population or patients' associations.

(iv) Who are the potential beneficiaries of the actions being assessed?
Details of the individuals who will benefit from the programme should be described and quantified (diseases/conditions, age, sex, socio-economic class, location, etc.).

(v) What are the standards being used for the appraisal?
The standards are the assessment criteria against which the appraisal will be conducted.

(vi) What is the economic viewpoint?
An economic assessment may be conducted from the viewpoint of various bodies or interest groups, e.g. health insurance organisations, society in general, government, politicians, manufacturers, a group of patients or their relatives, the international scientific community, or health professionals as opposed to patients and/or families.

(vii) Who conducts the appraisal and where do they work?
- Which professionals are already involved in using and/or improving the screening programme to be assessed?
- Which other professionals should be involved to cover future extensions of screening and/or to ensure that the appraisal has the maximum impact?
- What contracts are in place or are planned between these professionals and the body requesting an appraisal?

(viii) What are the different stages of the appraisal?
- What is its schedule?
- Is the timing good? (The answer will depend on the intended purpose and the aims being assessed.)

The appraisal of a screening programme may be carried out at different time points. This guide covers an *ex ante* evaluation, i.e. a study of the feasibility and impact of a planned programme.

(ix) What are the structures and budget for carrying through an appraisal of the screening programme?

(x) What are the limits of the appraisal?
The appraisal plan should prioritise the questions raised by the body making the request according to the expected consequences (decision-related or others) of the appraisal, the information available and the resources to be allocated to the appraisal. The plan should clearly state any points that are not being addressed.

(xi) How will the body requesting the appraisal make the results known?
If an appraisal is to be effective, there must be a policy for making the results known and implementing them.

Step 2. Search the literature

The aim of the literature search is to collect all the relevant published data which will be used to answer the problems raised by the proposed screening programme. It must be consistent with the rationale given and each question should have its own specific literature search.

We propose a standard search strategy. Databases, types of systematic search and keywords are given in Table 3. Further searches may be performed depending on the issues (Table 4). A search of the grey literature (not indexed in the databases) on French screening programmes is particularly important for an overview of the situation in France.

• Which are the available literature sources?

• *Electronic literature databases (list not exhaustive)*
- Medline (National Library of Medicine, United States)
- Embase (Elsevier, Netherlands)
- Pascal (CNRS-INIST, France)
- Cochrane Library (United Kingdom)
- BDSP (Public Health Database, Rennes)
- CODECS and NHS EED (The *Collège des économistes de la santé* and NHS Economic Evaluation Database databases).

• *Grey literature sources (not indexed in the databases)*
- for French programmes (Ministry of Health, CNAMTS, URCAM, URML, ORS, FAQSV, SUDOC (for theses) etc.)
- for foreign programmes (optional search): websites of foreign Ministries of Health or governmental or non-governmental bodies concerned with screening: National Health Service (NHS, United Kingdom), Australian Institute of Health (AIH), Santé Canada, Centers for Disease Control (CDC, United States), INAHTA (International Network of Agencies for Health Technology Assessment*).*

• Standardised literature search

The Medline, Embase and Pascal search strategy specifies the search terms used for each subject or type of study and the search period. These terms are taken from a thesaurus (eg MeSH descriptors for Medline and EMTREE for Embase) or from the title or abstract. They are combined in as many steps as required using the operators “AND”, “OR” or “NOT”. In Table 3, when the search field is not specified, it is the descriptor field.

A table should list steps and highlight results for:
- total number of articles retrieved
- number of articles analysed
- number of articles cited in the final list of references in the appraisal report.

Table 3. Literature search strategy

Study type/subject	Terms used	Search period
Existing guidelines		
STEP 1	keywords specific to the disease	
AND		
Step 2	*Guideline** OR *Practice guideline* OR *Health planning guideline* OR *Recommendation* [title] OR *Consensus development conference* OR *Consensus development conference, NIH* OR *Consensus conference* [title] OR *Consensus statement* [title]	
Meta-analyses, literature reviews		
STEP 1		
AND		
Step 3	*Meta analysis* OR *Review literature* OR *Literature review* OR *Systematic review* OR *Review effectiveness* [title]	
Screening		
STEP 1		
AND		
STEP 4	*Screening* OR *Mass screening* OR *Screen** [title]	
Epidemiological data		
STEP 1		
AND		
STEP 5	*Epidemiology* OR *Prevalence* OR *Incidence*	
Economic data		
STEP 1		
AND		
STEP 6	*Cost allocation* OR *Cost-benefit analysis* OR *Cost control* OR *Cost of illness* OR *Cost savings* OR *Costs and cost analysis* OR *Cost effectiveness* OR *Economic value of life* OR *Health care cost* OR *Health economics* OR *Economic aspect* OR *Pharmacoeconomics* OR *Cost(s)* OR *Economic(s)*	
Screening programmes		
STEP 1		
AND		
STEP 7	*Program Evaluation* OR *Public Health* OR *Health Priorities* OR *Health Planning* OR *Health Planning Guidelines* OR *Health Services Research* OR *Program** [title] OR *Campaign** [title]	
French articles		
STEP 8		
	Dépistage	
	Total number of references obtained **X**	
	Total number of articles analysed **Y**	
	Number of articles cited **Z**	

Table 4. Examples of additional searches

	Type of search
Diagnostic test	A search for guidelines, meta-analyses and systematic reviews should identify any reliable diagnostic tests. If no such tests are found, the search should be restricted to quality of diagnosis. (keywords specific to the disease) AND *Diagnosis* AND *Diagnosis, differential* OR *Diagnostic errors* OR *Controls, Quality* OR *Quality control, etc.*
Treatment	A search for guidelines, meta-analyses and systematic reviews should identify any effective treatment for the disease. If no such treatment is found, the search should be restricted to randomised controlled trials. (keywords specific to the disease) AND (*Drug therapy* OR *therapy* OR *Surgery, etc.)* AND *(Randomized controlled trial, etc.)*
Economic studies	This search concerns the cost of the disease. (keywords specific to the disease) AND (*Burden of disease* [free text] OR *Cost of illness* OR *Health Care Cost* OR *Coût de la maladie* [free text], *etc.*)
Natural history of the disease	(keywords specific to the disease) AND (*History, natural* OR *Natural history* [title], etc.).

Step 3. Set up a working group and a team of peer reviewers

- **Working group**

The working group may consist of up to 15 experts. All medical specialties concerned by the disease should be represented (with at least one public health physician, one epidemiologist and one economist in each working group). A biomedical engineer may also be needed to describe the French market for certain screening tests (the companies producing and marketing these tests may not be the same in different countries).

The working group's tasks are to critically review the draft report, provide further literature references, highlight any controversial issues, and give an opinion on practice.

- **Peer reviewers**

The team of peer reviewers should include other professionals to make it more representative. It may include representatives of patients' associations. The peer reviewers' tasks are to assess the readability of the report, to criticise its content and form, and to give an opinion on its conclusions.

- **Other contacts**

A meeting might be arranged with the companies who market a screening test and with representatives of user groups concerned by the disease.

III. Data collection and appraisal criteria

Screening is relevant when it helps to improve morbidity and mortality in a population. Relevance is judged by a list of criteria relating to:

- the nature of the disease to be screened for, screening test properties, diagnosis and treatment of the disease screened for,
- the efficacy and safety, economic implications, administrative arrangements and assessment of the screening programme.

The list of criteria (7) (shown below as items in boxes) is based on the list first produced by WHO in 1966 (8) (Annex 4). It has been expanded to include criteria from Canadian (9) and American (10) standards designed to improve screening efficacy and take account of any adverse effects on health. Ideally, all the criteria below should be met before implementing a screening programme. Reasons should be given if one or more of these criteria are not used in an appraisal.

An appraisal addresses the size of the target population, the resources needed for adequate coverage, any costs beyond current costs, and anticipated health gains. It should provide an estimate, based on full knowledge of the facts, of the benefits and priority to be given to the introduction of such a programme.

III.1. Collect data on the disease

- **The repercussions of the disease on the individual and on society should have been measured in terms of morbidity and/or mortality and socioeconomic impact.**
- **The epidemiology and natural history of the disease, including development from latent to declared disease, should be adequately understood.**
- **All the cost-effective primary prevention interventions should have been implemented as far as possible.**

• Have the repercussions of the disease on the individual and on society been measured?

There is no gauge to determine the impact on public health of the disease to be screened for. Impact can be measured from repercussions on individuals and on society (mortality, morbidity, socioeconomic impact).

- *For the individual*, the burden of the disease is given by the potential years of life lost, the cost of managing the disease, the degree of disability, pain and discomfort, and the impact on the family. Patients' quality of life may deteriorate. They may have financial problems in addition to the pain and suffering due to disease and the likelihood of premature death.
- *For society*, the burden may concern the community at large or just the patient's environment:
  - the burden for the community may consist of the mortality, morbidity and cost of managing the disease, and loss of productivity;
  - family and friends may incur financial loss and may undergo psychological and emotional trauma in addition to the distress they feel. The disease may have many social (stigmatisation of a group) and financial repercussions.

The size of the target population for the screening programme can be estimated by analysing the epidemiological data.

- **Do the epidemiology and natural history of the disease justify introducing a screening programme?**

The natural history of a disease is its spontaneous course without treatment. It is divided into four steps (11):

1. initial biological changes, which are generally undetectable;
2. first preclinical manifestation of abnormality; there are no clinical signs of disease yet, but the disease can be detected by appropriate tests;
3. clinical manifestations of the disease which make it possible to detect its presence and identify it;
4. outcome of the disease: recovery, complications, death.

In theory, treatment is more likely to be effective at a lower cost when it is given early during the course of the disease, at a time when the disease process may be reversible and complications have not yet occurred. The critical points are:

- the point when the disease can be detected and is not yet causing physical or mental deterioration (preclinical stage of disease);
- the point after which the disease will cause irreversible physical or mental deterioration;
- the stage during which treatment will counter the onset of the physical or mental deterioration, more effectively or with fewer consequences.

During the course of the disease to be screened for, there should therefore be a sufficiently long period without symptoms or during which symptoms are easily missed.

- **Primary prevention**

All the cost-effective primary prevention interventions should have been implemented as far as practicable.

III.2. Collect data on the test

- **There should be a simple, reliable, reproducible and valid screening test.**
- **The test should be acceptable to the population.**

- **The qualities of a screening test**

A screening test helps select from within the target population individuals who might have a given condition. It is applied systematically before any symptoms are evident and not according to whether the patient has any symptoms. The qualities of a screening test are given in Box 2.

A test that is not accepted by the public is likely to result in low participation and in poor compliance with the screening programme.

Box 2. Qualities of a screening test

Simple and easy to perform	This is especially true if the screening programme involves a large population. The test must be easily performed by a large number of doctors and technicians.
Reliable	The test result should correspond to the abnormality being looked for.
Reproducible	The test should give the same results when it is used on another occasion under the same conditions in the same subject, by other investigators or in other places
Valid (2)	The results should help differentiate individuals who may have the condition from those who do not.
Acceptable to the public	The test should be as non-invasive as possible, presenting no hazards. This is particularly important as screening is by definition intended for symptom-free individuals

• The performance of a screening test

There are two types of test properties:
- *sensitivity and specificity*, giving the intrinsic validity of a screening test
- *predictive values,* relating to the use of the test for a given population.

The performance of a screening test depends on both its intrinsic properties (sensitivity and specificity) and the type of population tested (particularly on disease prevalence). One needs to know:
- its *intrinsic performance* determined under experimental conditions
- its *extrinsic performance* determined in a given population in a screening situation.

A screening test concerns individuals who are apparently in good health, and it is carried out in population groups. In a screening situation, sensitivity often prevails over specificity. Many individuals suspected of having the condition are found, together with many false positives. A second, specific test is then used to confirm the diagnosis and eliminate false positives (see definitions in Box 3).

Box 3. Definitions of test properties and results

Sensitivity	Probability that a test is positive if the subject has the disease
Specificity	Probability that a test is negative if the subject does not have the disease
True positives	Positive results in subjects who have the disease
False positives	Positive results in subjects who do not have the disease
True negatives	Negative results in subjects who do not have the disease
False negatives	Negative results in subjects who have the disease
Positive predictive value (PVV)	Probability that a subject with a positive result is a true positive
Negative predictive value (NPV)	Probability that a subject with a negative result is a true negative

• Five questions to be asked when reviewing studies of the performance of a screening test

The performance of a screening test is determined in studies whose quality and validity must undergo critical review. Five questions should be asked.

1. Does the study have good internal validity (12)?

- The screening test should be compared with the most valid gold standard diagnostic test (the one which is best at categorising an individual as ill or not ill).
- The results for the screening and gold standard tests should be compared blind to ensure objectivity and avoid information bias. Failing this, a rank effect should be avoided.
- To avoid "verification bias", conducting the screening test result should not be a reason for not performing the gold standard test in all patients. The assessment of the intrinsic validity (sensitivity and specificity) of the screening test would be falsified if its results were to influence the decision to carry out the gold standard test.
- The study should state how the test should be performed (materials and methods). This includes patient preparation (diet, medicines to be taken or avoided, precautions to be observed after the test), the test itself (technique used, whether the test is invasive and/or painful), and the method used to analyse and interpret the results (13).

2. Is the intrinsic performance of the test given?

Sensitivity and specificity should have been correctly calculated in the article or it should be possible to calculate them from the data provided. Confidence intervals should be given. How the threshold of normality was determined should be described. Reproducibility depends on the accuracy of a measure. It may be given by the degree of intra- or inter-observer agreement in *ad hoc* tests.

3. Is information available on the prevalence of the condition in the test population?

The basic question that a clinician has to ask when a subject assumed to be healthy so far has a positive (or negative) screen test is: what is the likelihood that the patient does (or does not) have the disease? The probability depends on the sensitivity or specificity of the test and on the prevalence of the disease in the given population.

The positive predictive value of a screening test (probability that a subject with a positive result is a true positive) decreases as disease prevalence decreases, even if sensitivity and specificity are high. If disease prevalence is low, the probability that a subject identified as positive actually has the disease may be relatively low. In this case, further tests are needed to confirm that the patient has the disease.

4. Have the consequences of false-positives and false-negatives been assessed?

The harm done by falsely identifying subjects as negative or positive must be assessed.

- Patients identified as false-negatives may have a false feeling of security that may make them overlook signs and symptoms. This could delay diagnosis (12,14,15) and lead to loss of confidence in the health system.
- Patients identified as false positives may suffer anxiety through fear of a serious diagnosis, may undergo invasive interventions to confirm or eliminate the diagnosis, be treated for a non-existent condition and possibly suffer from the iatrogenic effects of this treatment. Other disadvantages are the costs borne by society (cost of lack of quality in the healthcare system) and collective resources that have been wasted and used to the detriment of other interventions.

5. Will the screening test modify management? Has its use been assessed outside centres of excellence?

Whether a screening test can contribute to improving patient management and, if possible, disease outcome has to be considered. It is also necessary to consider whether results similar to those in study participants can be achieved in patients in clinical practice (16). Criteria for establishing whether they have a sufficient number of properties in common are demographic characteristics, level of care (primary, secondary or tertiary), type of patient, and patient selection. If there are enough properties in common, and if the reader's experience of practice is similar to that described in the article, the screening test may be expected to be as effective in that practice as in the published study.

In 2003, the STARD committee (Standards for Reporting of Diagnosis Accuracy) proposed a checklist for writing a critical and exhaustive review of studies on diagnostic test performance (17) (Annex 5). This list is also a useful tool to assess the performance of a screening test.

III.3. Collect data on diagnosis

There should be an agreed policy within the scientific community on further diagnostic investigation of individuals with a positive test result and on the choices available to those individuals.

Care procedures and networks need to be defined in terms of accessibility, availability and technical requirements. Recall procedures are needed for individuals missing appointments.

III.4. Collect data on the intervention

"Intervention" means treatment, a preventive measure, or information felt to be important for the individual with the disease.

- **There should be effective intervention for patients identified early, with evidence of earlier intervention leading to better outcomes than later intervention.**
- **There should be agreed evidence based policies on which individuals are likely to benefit from the intervention and on the appropriate intervention to be offered.**

There should be evidence that early intervention gives better results than late intervention. Some treatments, however, delay the onset of complications rather than cure patients.

III.5. Assess the effectiveness and safety of the programme

The effectiveness and safety of the screening programme should be assessed in comparison to all alternative strategies.

- **There should be evidence from high-quality randomised controlled trials or from an international consensus that the screening programme reduces mortality or morbidity.**
- **The benefit from the screening programme should outweigh the harm caused by the test, diagnostic procedures and interventions.**

• Criteria used to assess the effectiveness of a screening programme

The criterion should reflect the state of health of individuals undergoing screening (mortality, morbidity). Criteria such as laboratory values, radiological criteria, or other endpoints often used in clinical trials are not appropriate for assessing a screening programme designed for symptom-free individuals who consider themselves in good health.

The ideal criterion is the disease mortality or morbidity rate. A reduction in mortality indicates that disease-related deaths have been prevented or, at least, have been detected after the follow-up period. A reduced incidence of the disease is also a relevant criterion when a "pre-disease state" is detected (cervical cancer, bowel cancer, hypertension).

The level of take-up of a screening test is an indicator of test acceptance by patients and doctors. It can be used to assess the potential effectiveness of a screening programme. Below a certain take-up rate, the programme is no longer effective from a community standpoint. Take-up rate is available soon after the study begins. A very low rate would make it pointless to continue the study, but a high rate does not necessarily mean that screening is effective. Take-up is an easily obtained indirect indicator which can be used as a secondary criterion (18).

• What NOT to do when assessing the effectiveness of a programme

It is often stated that treatment is more effective when given early in the course of a disease and therefore that the disease was diagnosed at an early stage. This statement is often based on faulty reasoning (19).

Do NOT compare mortality or morbidity before and after screening

This type of comparison has the same disadvantages as historic controls. Screening is very rarely the only difference between the two periods studied.

Do NOT compare survival rates

It is often thought that the survival of patients with disease detected during screening is better than that of patients diagnosed outside screening. This is not true for the four reasons (even if the date of death has not changed by an iota) given in Box 4.

• Look for randomised controlled trials (RCTs) as evidence of effectiveness

The first effects of a screening programme are:
- an increased incidence of the disease being screened for (because of early diagnosis and particularly overdiagnosis)
- diagnosis of a greater proportion of individuals with less advanced disease
- increased survival (due to earlier diagnosis, selection of cases with a better prognosis, overdiagnosis and selection of the population screened).

Measurement of mortality in the whole target population can eliminate the first 3 types of bias in Box 4 (20). Only an RCT, however, will avoid the population selection bias of comparative studies, e.g. case-control studies, which compare the history of patients who have died or who have advanced disease with that of controls from the general population (19).

Because in RCTs individuals from the target population are randomly allocated to a screened or control group, the proportion of patients with rapid- and slow-developing disease is likely to be similar in both these groups. The aim is to demonstrate a reduction in disease

outcomes, notably in mortality, by comparing disease-related mortality rates and not survival curves (see Box 4) (19). Each patient, whichever group they belong to, receives the intervention regarded as the most appropriate as soon as the diagnosis has been confirmed. Quality and duration of follow-up are the same in both groups (12).

Box 4. Why one cannot compare survival rates of patients diagnosed within and outside of a screening programme

1. Zero-time shift or lead-time bias

Survival seems to be longer in patients diagnosed earlier. This bias is called zero-time shift or lead-time bias. It corresponds to the time interval from when the disease was detected through screening to when it would have been diagnosed without screening. Early diagnosis does not therefore postpone death but leads to longer disease. It is essential that the increase in survival is longer than the time gained through screening (11). Mortality is the only criterion that can demonstrate increased survival resulting from improved prognosis (12).

2. Selection of patients with a better prognosis (length bias)

This bias is due to different progression rates of tumours and other serious diseases. Diseases progress at very different rates in different patients, for reasons that are generally not well understood (concomitant disease, nutritional deficiencies, genetics, etc). The progression rate tends to be reflected in the natural history of the disease. Disease progresses longer in patients with a longer preclinical phase. The longer this phase, the greater the likelihood of diagnosing the disease early. Screening is therefore likely to take place during the preclinical phase of a disease with a slow rather than a fast rate of progression. Diseases with a long survival are at an advantage (11). Any comparison of survival for diseases detected through screening and diagnosed clinically will thus suffer from bias. The bias is greatest when screening is introduced into a population and decreases as the number of rounds of tests increases (12).

3. Overdiagnosis

Screening may reveal diseases which would not have progressed and would never otherwise have been diagnosed. Better survival in these cases does not mean that screening is effective but indicates a selection bias in favour of forms with a good prognosis (11).

4. The healthy volunteer effect or selection by screening

Screening is offered to the whole of the target population but not all individuals will take up the offer. Those who accept the constraints of periodic screening may have a better baseline state of health and a lower incidence of disease than those who refuse and with whom they are compared. Results in volunteers cannot therefore be generalised to other groups. This bias disappears if a high proportion of the target population takes up screening (12).

Recent RCTS tend to randomise subjects not individually but in clusters. The clusters may be hospitals, doctors, families or villages, i.e. social units whose members are considered to be interdependent. The intervention is thus applied at a level above the individual. Cluster randomisation is justified if contamination between the groups to be compared is suspected. Contamination would lead to a bias in the estimated effect and would mean that the statistical hypotheses based on independence do not apply. Cluster randomisation has two implications for study design (22):

- The calculation of the number of subjects needed must take account of correlations within the data. Sample size must be increased accordingly. The results observed in 2 subjects in the same cluster would tend to be more “similar” than results in 2 subjects from 2 different (and therefore independent) clusters.

- In cluster randomisation, randomisation usually precedes subject recruitment. Clusters (doctors, hospitals etc) are randomised first and each “cluster manager” is then responsible for including subjects. When the trial is an open study (as for an intervention such as a patient management programme, a diet, etc), subjects are recruited in full knowledge of the arm to which they will be allocated. This is potentially a source of bias, and a retrospective comparison of the groups formed by randomisation then becomes even more important.

• Be certain that follow-up is long enough and sample size appropriate

Follow-up should be long enough to measure the endpoint chosen to assess the screening programme's outcome. The end-point may be the absence of an event and not its presence, i.e. the disease does not cause complications or causes fewer complications with regard to morbidity or mortality. Extended follow-up is needed to demonstrate later onset of a complication in the screened *versus* control group, especially when the disease progresses slowly. The time needed might even compromise the feasibility of an RCT (18). Moreover, advances in tests and interventions may render the results obsolete several years after the start of the trial (18).

Population studies are needed to judge the effectiveness of a screening programme, but choosing sample size is difficult. The choice is often based on theoretical factors or on publications that can lead to an estimate of the maximum expected effect. Sample size has consequences in terms of feasibility, organisation, costs, and maintaining comparability of groups, with the problem of subjects lost to follow-up.

• Use a checklist to assess the screening programme

A checklist to assess a screening programme has been published by Earle and Hébert (23):

1. What were the objectives of the study?
2. Were the candidates well described?
3. Is the study randomised?
4. How well was the randomisation done?
5. Was the screening procedure well described?
6. Was the follow-up > 80%?
7. How were non-compliers analysed?
8. What outcomes were measured?
9. If a positive study, were both clinical and statistical significance considered? If a negative study, was the power assessed?

IV. Economic assessment of a screening programme

A screening programme is warranted when it is cost-effective compared to no screening or individual screening, and when it is preferred to another health initiative by the financing body.

An assessment is not restricted to medical criteria alone. Economic assessment is included in the WHO criteria (8) and should be included in any appraisal of a screening programme.

IV.1. What is the part played by an economic assessment?

Economic assessment involves defining the investment thresholds beyond which the community may consider that resources are no longer being spent usefully and would achieve greater well-being (e.g. save more lives or better satisfy needs) if allocated elsewhere (opportunity cost). An economic evaluation helps to answer two key questions (3,24):

1. *What are the priorities in the use of healthcare resources?*

The benefit of a screening programme should be analysed in terms of overall healthcare resource allocation and take account of the scale of the health problem (incidence and severity of the disease, proportion of the population concerned), the costs of screening, its effectiveness, costs that it would avoid, and possible alternative solutions.

2. *What type of screening programme (organised or opportunistic, centralised or decentralised) and what type of test should be used?*

The decision depends on examining the efficiency of each strategy and selecting the dominant ones, i.e. no other strategies are both less expensive and more effective. Questions needing answers are:

- What are the possible strategies in the light of the programme's aims (age groups and target population groups, frequency of screening)?
- What results will be achieved by each of these strategies?
- What resources does each of these strategies require?
- Which of these strategies are technically effective and have a cost-effectiveness acceptable to decision-makers? A decision to introduce a screening programme means agreeing to pay substantial costs for a hypothetical extra "statistical" year of life in the target population. According to some international institutions, a medical intervention is efficient when cost ratios per life-year in good health gained are less than GDP per capita. In Europe, this is approximately EUR 25 000, which is about half the threshold value often used in published data (24).

Like most medical interventions, screening is subject to decreasing returns. At constant sensitivity and specificity, the positive predictive value (PPV) of a test decreases rapidly as screening is extended to population groups with a lower prevalence of disease or of a risk factor. As screening is extended, more tests have to be carried out to detect an extra case (25). In economics, this is called marginalist reasoning. In the public sphere, screening is justified up to the point where its marginal cost (cost related to detection of one additional case) becomes equal to its marginal benefit (complication or additional death avoided, additional year of life gained). Beyond this point, resources would definitely be better used elsewhere (opportunity cost).

An economic evaluation can be carried out:
- when deciding to introduce screening (forecasts of anticipated costs and results, establishing the conditions under which it will be introduced, etc) (this is when evaluations are performed by ANAES)
- when decisions on managing the programme are taken
- when the decision is taken to repeat the operation (actual costs and results).

Moatti (25) has described how economics can help to answer the questions raised by an appraisal of a screening programme (Box 5).

Box 5. Possible contributions of an economic assessment to appraisals of a screening programme according to Moatti (25)

1. Determining conditions of use, frequency and the "ideal" distribution of screening policies

- What is the justification for mass screening?
- Which population groups (prevalence, risk factors, age range) should be offered screening?
- What procedures (frequency of screening and test sequences) should be implemented?

2. Explaining the consequences of a decision by extending the judgment criteria to include

a. Medical efficacy.
b. Analysis of financial flows.
c. Harmful effects of screening (anxiety, stress and iatrogenic risks associated with false positives; delayed treatment in the case of false negatives.
d. Equity.

3. Describing how actual programmes fulfil or fail to fulfil their stated aims; identifying any sub-optimal elements so they can be remedied

a. Assessment of organised screening programmes.
b. Comparison of tools or organisations (how individuals will be invited to attend, procedures to ensure coordination between those involved, the degree to which technical resources will be decentralised or mobile, etc.).
c. Acceptance level.

4. Determining the financial and organisational constraints and inducements most likely to promote a more rational use of resources

a. Better effectiveness and allocation of resources for organised screening programmes than for haphazard distribution of screening tests and early diagnosis.
b. Organisational and financing procedures compatible with institutional and sociological constraints re most likely to ensure that target groups have access to screening.

5. Analysing the obstacles encountered by screening programmes to encourage take-up and ensure access by certain categories of individuals

a. Completing sociocultural and sociological approaches.
b. Analysing behaviour in the face of risk.

6. Determining the macroeconomic and macro-social consequences of introducing screening to predict very common diseases (e.g. cancer).

a. Impact on current social insurance systems.
b. Regulatory measures (self-regulation by health professionals, adaptation of the health insurance system, etc.).

IV.2. Methods of economic assessment

The methods are used to decide between alternative strategies, the decision resting on a number of criteria. Each method has its field of application and limitations. Each differs in the way consequences are measured (Table 4). All methods have been the subject of international (26-28) and French (24) guidelines.

Table 4. Types of economic assessment studies

	Measure of costs	Identification of consequences	Measure of consequences	Variable
Cost minimisation	Monetary	Same consequences for all options	None	$C_1 - C_2$
Cost-effectiveness	Monetary	One-dimensional indicator of effectiveness	Physical unit	$\frac{C_1 - C_2}{E_1 - E_2}$
Cost-utility	Monetary	Multidimensional indicator of effectiveness	QALY	$\frac{C_1 - C_2}{QALY_1 - QALY_2}$
Cost-benefit	Monetary	Multidimensional indicator of effectiveness	Monetary	$(C_1-B_1)-(C_2-B_2)$

QALY: *Quality Adjusted Life Year*

Cost minimisation studies are used when clinical trials have shown that the results of competing policies are equivalent. The decision is then based on lowest cost alone.

Cost-effectiveness studies study both costs and results, provided that results are expressed by a clinical indicator or an objective indicator of state of health (number of cases screened, complications avoided, life-years gained, etc.) (24).

Cost-utility studies measure health results not in quantitative terms alone, but by introducing the concept of quality of life. In a cost-utility analysis, derived from the theory of utility, a single indicator summarises quantitative (gain in life expectancy) and qualitative (reduced morbidity, improved or worsened quality of life) information. A single criterion is used to compare health initiatives.

Cost-benefit studies help determine whether a new health policy offers a clear benefit for society. They compare cost with willingness to pay, i.e. the amount of money that the community is prepared to pay to obtain an additional unit of health. A cost-benefit analysis differs from a cost-effectiveness analysis in that all costs and consequences are expressed in monetary terms.

The type of study will depend on the problem and the body making the request. Cost-utility and cost-benefit studies are rarely used to appraise screening programmes although taking utility, i.e. the satisfaction of individuals, into account would be a worthwhile option.

IV.3. Analysing economic studies

The method proposed by ANAES involves making the best use of existing data, in particular published clinical and economic studies as well as health insurance data, rather than carrying out studies "in the field". The method is based on a critical literature review. Drummond (26) has made a checklist of the properties of well designed studies (see Annex 6). Its use should quickly highlight the strengths and weaknesses of a study.

Economic studies must satisfy two types of criteria:

(i) Criteria of internal validity (see Annex 6 and items in Section VI.4) which ensure:
- *credibility of results*. This assumes transparency of methods and data sources, and the use of sensitivity analyses;
- *study design quality*. The methods used should be appropriate for the intended aims;
- *comparability of results*. The choice of indicators of costs and results must be described in detail.

(ii) Criteria of external validity which concern:
- *clinical results*. It is generally accepted that clinical results can be transposed from one country to another, or from a group of countries to a single country, if criteria defining the population and the organisation of the healthcare system are met.
- *epidemiological data*. These may vary more widely among populations.
- *health services*. The cost of medical procedures varies among countries. A screening programme must be adapted to local conditions. Consistent policies must be adopted (29).
    - There are differences in patient management and therefore in the use of resources, and differences in prices and fees for materials and services.
    - If prevention behaviours (including differences in participation levels) are to be transposed from other countries to France, cultural differences and population-specific factors need to be taken into account.

Epidemiological or economic data cannot therefore be transposed from one country to another without making the appropriate checks beforehand. When an economic assessment uses international data or national data for another country, the relevance of the data (clinical, epidemiological or economic) has to be confirmed (24).

IV.4. The six tasks in an economic assessment

1. Specify management segments

The segments (screening, diagnosis and treatment) to be used as a basis for the economic assessment must be defined. Diagnosis costs may need to be included in screening costs because of resources used in relation to false-positive diagnoses.

2. State the viewpoint

The viewpoint and costs to be taken into account depend on the persons who request the study and/or for whom the results are intended (24). The choice of viewpoint is important. A programme that appears poor from one perspective may be much better from another. Possible viewpoints are: patients, an institution, the target group for the programme, the Ministry of Health budget, the government budget (Ministry of Health and other ministries) and, finally, a societal viewpoint (26).

As the aim of the economic assessment is to provide a public health policy decision-making aid, the viewpoint should ideally be a societal viewpoint. In France, this may embrace collective benefit, a general public health perspective or issues of equity between groups and between generations. A societal viewpoint is rare in economic assessments as the costs are difficult to demonstrate and calculate.

3. Choose a reference policy

The proposed screening programme should be compared with one or several reference policies chosen from an exhaustive list of policies (24). The external validity of an economic assessment is jeopardised if the reference policy is not relevant (26).

- If there is already a screening programme, current screening practice will be the reference policy.
- If there is no screening programme, the reference policy will be absence of screening, i.e. current practice for diagnosing and treating symptomatic patients.

In practice, it may be useful to compare all available screening policies with that of no screening, i.e. with a common reference policy. This will tell whether the current screening policy is open to criticism or questionable.

4. Highlight types of cost

Economic assessment traditionally examined 3 types of costs (direct costs, indirect costs and intangibles) but nowadays examines only the first two. Intangibles which reflect human and psychological costs tend to be analysed in quality of life measures (24).

- If the assessment is limited to intermediate screening efficacy, i.e. cost per case detected, there are probably no indirect costs (no loss of production nor loss of productivity).
- If the end-point is final efficacy from a society viewpoint, indirect costs must be taken into account.

Irrespective of the type of costs analysed, it is important to reason in terms of marginal cost rather than average cost (Box 6). It is the only option that is valid as it defines the optimum thresholds beyond which the community may consider that resources are no longer being spent usefully (30).

Box 6. Definitions of costs

Average cost	Cost of production per unit or unit cost; equal to the total cost divided by the quantity produced.
Marginal cost	Cost of the last unit produced; corresponds to the variable costs involved in increasing production by one unit, as the fixed costs have been distributed over units already produced.
Direct costs	Total of resources consumed that are directly attributable to screening; comprises medical costs and non-medical costs (Fig. 3).
Indirect costs	Mainly loss of production, productivity and human life; costs due to screening and related to the productivity of active patients and, consequently, national wealth.

Direct medical costs consist of human resources (time spent by health professionals, consultations) and material resources (cost of materials, tests and investigations, cost of immediate treatment if available). An economic assessment can rarely be limited to the act of screening itself as, in practice, screening generates medical follow-up or diagnosis costs.

Non-medical direct costs include the cost of providing information for the population. This is an important but not unique factor in the success of screening. The cost of information campaigns (documents, posters, press, conferences, meetings and study days, audiovisual materials) and the cost of postal invitations (printing of documents and mailings) must be included. Non-medical direct costs include:

- running expenses (premises, office equipment, telephone, postage, etc) in the case of centralised screening (e.g. at a hospital, in screening centres etc);
- computer processing costs (purchase of equipment, software, operation, data entry) (31).

Finally, some non-medical direct costs may be incurred by patients or those around them (cost of transport or childcare, working or leisure time) (32).

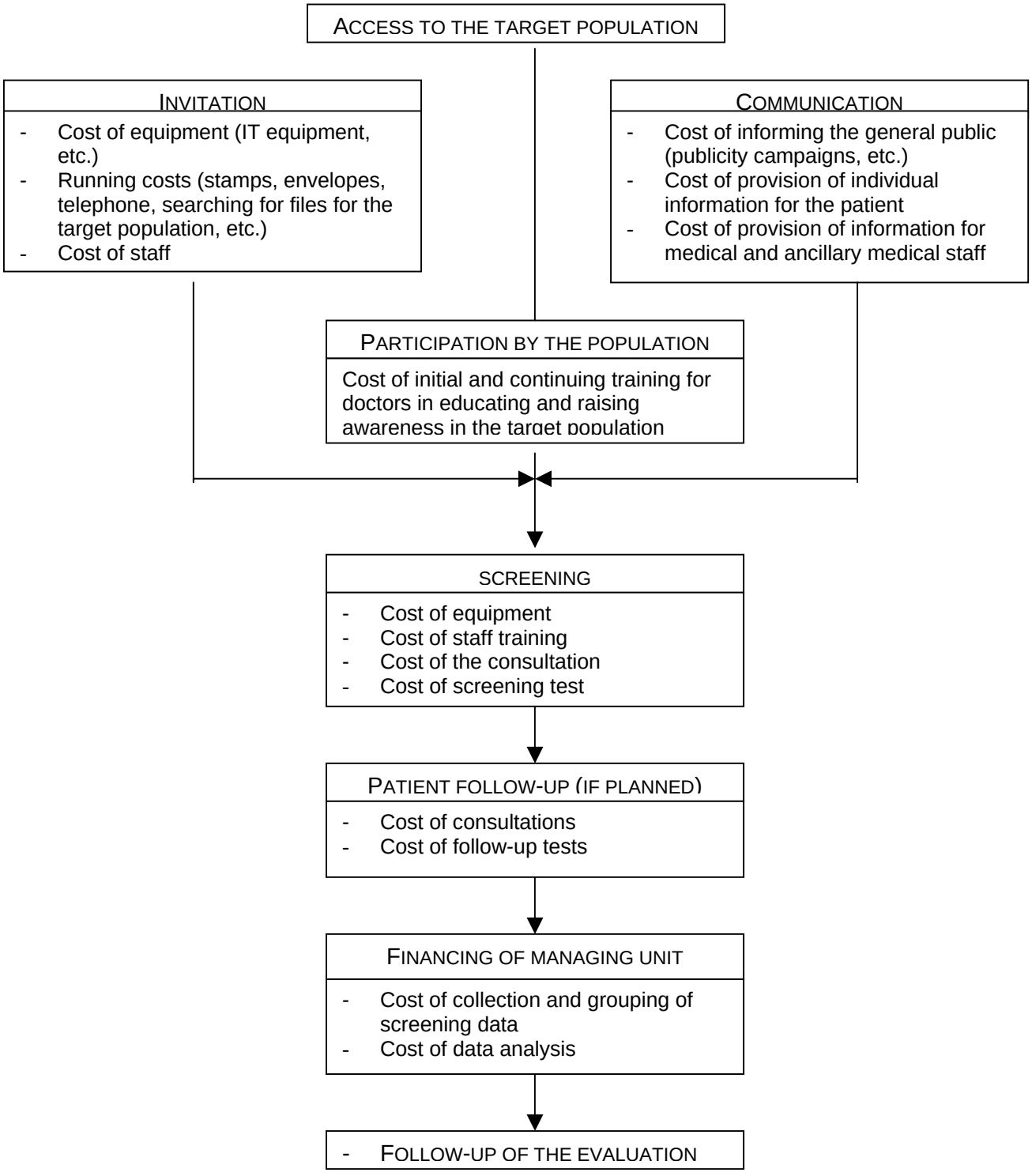

Figure 3. Valuation of steps in a screening programme (direct costs) after Brown, 1998 (32)

5. Justify the time scale (discounting, time taken to analyse results)

The costs and benefits of screening are spread out over time and do not coincide. Costs should be discounted in economic analyses to take account of time (32,33). However, there is no consensus among health economists on this (24).

Discounting involves asking what a benefit or a cost would be on a date t_1 compared to the benefit or cost arising now, at t_0, at a constant euro value. Discounting tends to increase the costs of screening (costs defined at t_0 and benefits arising at t_1) compared to the costs of a curative action (costs and benefits defined at t_1),

Another controversial question is whether life-years gained can be discounted in the same way as cash flow can. Some economists feel that, unlike monetary revenues, life-years gained cannot be reinvested and deferred to the future (33).

The chosen discount rate depends on the hypotheses relative to the decision-maker's rationale and to the economic environment. For an international comparison, the *Collège des économistes de la santé* (24) recommends using the rates 0 (calculation without discounting), 3 and 5%. A sensitivity analysis should be conducted systematically on discount rates to test the robustness of the conclusions.

6. Verify robustness of results by a sensitivity analysis

A sensitivity analysis examines how systematically modifying key hypotheses or factors affects the results of the economic assessment (34). Some factors have greater impact than others on an economic assessment. These are:
- incidence of the condition in the population
- clinical effectiveness of screening
- performance of the diagnostic test (sensitivity, specificity and PPV)
- coverage
- acceptability of the tests used and participation in the programme (see section on the programme). Moatti (25) states that policies with the same level of acceptability must be compared in any appraisal
- costs
- discount rate.

A sensitivity analysis helps to overcome part of the uncertainty due to variations in data (degree of compliance with policies) and in the population sample under study (e.g. demographic changes), when data are extrapolated from one environment to another (34). It tests the robustness of results and determines how far variations affect the hierarchy of cost-effectiveness ratios (25).

The economic framework required and the components of the economic analysis are summarised in Table 5.

IV.5. Taking account of the social dimension of screening

Equity means achieving effective and fair treatment of individuals (36). According to the *Collège des économistes de la santé* (24), equity issues should be handled by taking relevant discriminatory factors into account when analysing the results of an economic assessment, i.e. by considering the variations in the distribution of costs and consequences of screening by group. Equity issues are:
- access to screening dependent on patients' income
- defining a target population at the expense of another
- a personal and socially controversial view of a survival indicator (a gain of one life-year is of identical value for all individuals throughout the life-span) (25).

Table 5. Economic framework required

Criteria to be satisfied	Specific to screening
1. Statement of the standpoint and question posed	The chosen standpoint is usually that of the financing body or society, as the action in question will use community resources and be aimed at part of the population.
2. Specific and explicit description of the clinical foundations for the study	Information is required about prevalence, screening tests used, thresholds used to define positive or negative patients, and rates of compliance with the protocol.
3. Description of the study population	The demographic characteristics of the population analysed should be given, so that it is possible to see how far studies' conclusions can be transposed and generalised to other populations.
4. Comparative analysis of policies	Each of the options should be stated clearly. The reader should have enough information about the different stages involved in each policy to decide whether all costs and consequences have been included.
5. Choice of most appropriate evaluation method and most relevant efficacy endpoint	The rationale should concern the final efficacy criterion (life-years gained, complications avoided, etc.).
6. Details of and justification for costs given and how they were valued	Costs should correspond to the point of view adopted and should be expressed in appropriate physical units. Direct medical costs should be identified, as well as other types of costs such as those related to introducing a screening campaign (administrative costs). In the analysis, it is important to prefer marginal costs to average costs.
7. All the consequences of the policies implemented taken into account and measured correctly	Measurement of the consequences of each policy studied will depend on the type of economic evaluation. If an economic analysis of screening is based on a cost-utility study, it is necessary to understand exactly how the result was adjusted to take account of the effects of implemented actions on quality of life.
8. Time taken into account	The long-term consequences of costs and health results should be estimated. The discount rate used should be justified.
9. Results take account of added benefit and added costs	An attempt should be made to determine whether there is additional benefit and whether this compensates for the additional cost generated by screening.
10. The analysis of uncertainty includes a sensitivity analysis and discussion of dubious methodological and clinical hypotheses	The prevalence rate for the disease and parameters of sensitivity and specificity of tests need to be varied to determine the impact on the preferred policy.

After Drummond et al., 1998 (26), CES, 2003 (24), Provenzale, 1996 (35)

V. Setting-up a screening programme

This section raises the issues relating to the setting-up of a screening programme but will not provide specific answers.

- **There should be a plan for managing and monitoring the screening programme and a set of quality assurance standards recognised by the medical community.**
- **Adequate investment in staff and equipment for carrying out the test, diagnosis, treatment and administration of the programme should be available before the screening programme begins.**
- **All other options for managing the condition should have been considered (e.g. improving global management).**
- **Screening should be a continuous activity, not a once and for all operation.**
- **For optimum participation by the target population, the best information possible should be widely diffused. Awareness programmes should be organised for both the target population and health professionals (3).**
- **Lack of information on positive and negative aspects of screening is not ethically acceptable, and it compromises the autonomy of the individual (3).**
- **If systematic screening can be offered, to ensure equity of access to screening, individuals should remain free to accept or refuse the test. Consent must be obtained after the patient has been informed about the advantages and disadvantages of screening.**

V.1. Prerequisites to setting-up a screening programme

As mentioned above, mass screening can only be recommended if:

- it avoids a significant proportion of new cases of the condition through detection and cure of a "pre-disease" state;
- it leads to a reduction in the mortality rate for the condition in the target population, through better treatment of the disease in detected cases;
- the negative effects are limited to a minimum; the pros and cons in terms of social and economic health costs have to be balanced.

When these conditions are met, it is necessary to ensure that screening will be part of an organised and planned programme. The set-up should guarantee benefits and limit disadvantages while keeping costs within available resources.

V.2. A case-by-case set-up: example of a cancer screening programme

The set-up will depend on whether the screening programme is aimed at a chronic disease such as cancer, an infectious disease (HIV) or a genetic abnormality (prenatal screening). As an example, a cancer screening programme should include the following (37):

Procedures for inviting the population to take part (information and encouragement)
If resources are limited, any decision to introduce a screening programme will mean that the allocated resources cannot be spent on other public health actions. Compliance by the population with the programme determines whether it is profitable. This involves obtaining the maximum response rate to the invitation from the target population, for a given recruitment initiative, and obtaining satisfactory compliance throughout the protocol. Patients lost to follow-up during screening entail a waste of collective resources as the expenses incurred by the first tests will not ultimately contribute to a diagnosis.

Procedures for informing the population should be defined. Written documents should be sent to individuals to help them understand the aims and procedures of the screening offer. They will be able to refer to them later and/or to discuss them with a health professional. The information in these documents and their production should be of high quality (38). If take-up is too low, the benefit of screening may be questionable as participating individuals will tend to belong to a population more concerned with its health. These individuals are likely to have already undergone investigations or to be under medical surveillance. This makes it less essential for them to have the test. Low take-up levels sometimes correspond to participation by a low-risk population only (39).

Screening policies should encourage information acquisition by the individual; individual benefit should be the priority (40). Not until the information needs for a satisfactory participation rate have been identified can the issue of whether screening is appropriate be addressed. Guides on information provision for individual candidates are needed but are not the subject of this report.

Quality control of tests, both technical performance and interpretation. This may result in the introduction of training and permanent monitoring.

Coordination between partners (general practitioners and specialists, ancillary medical staff (pharmacist, nurses). Their job is to inform and encourage candidates, technicians and operators.

Continuity of recruitment. Recruitment is not a one-off procedure. Screening "campaigns" or “weeks” may persuade individuals to submit to a test once, but the full benefit from follow-up is missing. A single test is of limited benefit for two reasons:
- in most cases, only a small proportion of the population will be screened;
- screening detects individuals with the disease at a given point in time but does not impact on the future incidence of the disease.

Continuous screening has major advantages. It will become more effective and less costly with time, and will become part of available healthcare services as a matter of course (8).

V.3. The pilot study

Before a screening programme is introduced into a large population, a pilot study should test the different components of the set-up within the local healthcare system.

V.4. Screening: just one way of controlling a disease

All other options for managing the condition (e.g. improving treatment) should have been considered. Screening is only one way of controlling disease. It should be part of a global action to reduce the burden of disease for the individual or the community through socioeconomic or environmental measures, health education and improvements in existing healthcare and prevention systems (3).

VI. Follow-up and appraisal of a screening programme

A limited number of the appraisal criteria and indicators should be validated; they should be chosen at the design stage and be based on the results of the literature review or the opinion of a panel of experts (1).

The data needed to appraise the screening programme proposal should be defined early on. The data will be used to judge the programme's feasibility (participation, compliance, number of cases detected, false positives, false negatives), costs (direct and indirect) and results (incidence, mortality).

VII. How appropriate is economic modelling

Simulation models are particularly useful to assess a screening programme and satisfy WHO criteria because field studies may not be possible for methodological or practical reasons, particularly when the disease has a long natural history. Moreover, assessment must be multidimensional as screening raises medical and psychological issues for the individual and economic issues for society.

VII.1. Definition of modelling

A model is a representation of the reality of an observed phenomenon or phenomena. Modelling is an analytical technique designed to simulate the impact of one or more variables on expected results (24). The model may range from the very simple to the highly complex. It must satisfy two competing constraints (41):

- It should be representative, i.e. capable of representing the real situation;
- It should be applicable, i.e. compatible with data collected about the real situation.

The more representative and therefore complex a model, the less applicable it is because of the level of detail required of the data.

VII.2. Aims of modelling

- **Aim 1. Summarise information from various sources**

One way of measuring the effects of an intervention that has never undergone a complete assessment is to use mathematical models, including decision trees. Models integrate information from a number of sources and represent interactions among variables (clinical decision analysis). A model will incorporate key factors that determine the success of a screening programme, i.e. the results of partial epidemiological studies: feasibility and acceptability of the screening procedure, the validity of the screening tests, the efficacy of treatments at each disease stage, and economic impact (clinical and economic decision analysis) (18).

- **Aim 2. Take account of uncertainty in screening procedures**

If there is considerable uncertainty about the effectiveness of screening and the most appropriate procedures, modelling can help to establish the minimum conditions and hypotheses to be satisfied to justify large-scale testing of a screening programme (25).

- **Aim 3. Take account of uncertainty in variables**

Models can take into account changes in the efficacy of tests and treatments, and also the characteristics of the target population. Results can thus be transposed to other populations. Moreover, a time factor can be introduced. This is important when screening for a disease with consequences that only occur a number of years later. In all cases, sufficient data must be available on the key points that determine the efficacy of the screening programme (18).

VII.3. Modelling techniques

- **Data availability**

Model use is subject to information availability. Two types of situation are encountered (41):

(i) *Data on the variable are available*. A critical review is made of the quality of data sources and of the precision of the estimated value for the variable. Ideally, data should be confirmed by 2 or more independent studies but this is rare. Information is often limited to a single study or to several studies with conflicting results. This does not preclude a decision analysis which is designed specifically to take account of and reduce sources of uncertainty.

(ii) *Data are lacking.* Lack of information on one or more of the model's variables should not be a reason for excluding these variables from the analysis or for not performing an analysis. The choice of variables to be included depends largely on a detailed description of the policies under study and of the key events in the model. The quality of a decision analysis depends more on the scientific relevance of the choices made about the policies than on whether the clinical, epidemiological and economic data required to carry out the analysis are available.

- **Types of approach**

Modelling techniques may be either determinist or stochastic (24). The most commonly used models in the field of health are decision trees and Markov models.

Determinist models: In a determinist approach, the model is based on estimates which are usually derived from published data (24).

- *A decision tree* is a schematic representation of the consequences of a treatment decision (Fig. 4).
- *Markov models* are built like decision trees but incorporate the concept of time in the form of a cycle (24). They are used for chronic diseases that require assessment over the long term.
- The *Monte Carlo simulation method* simulates a theoretical cohort of individuals; each chance event is simulated for each individual in the cohort (24).

Stochastic models: Unlike the determinist approach, the stochastic approach is based on individual observations or raw data, e.g. data from clinical trials, hospital records, or medical resource use files. Uncertainty is measured by observing statistical distributions of the variables of the model (24).

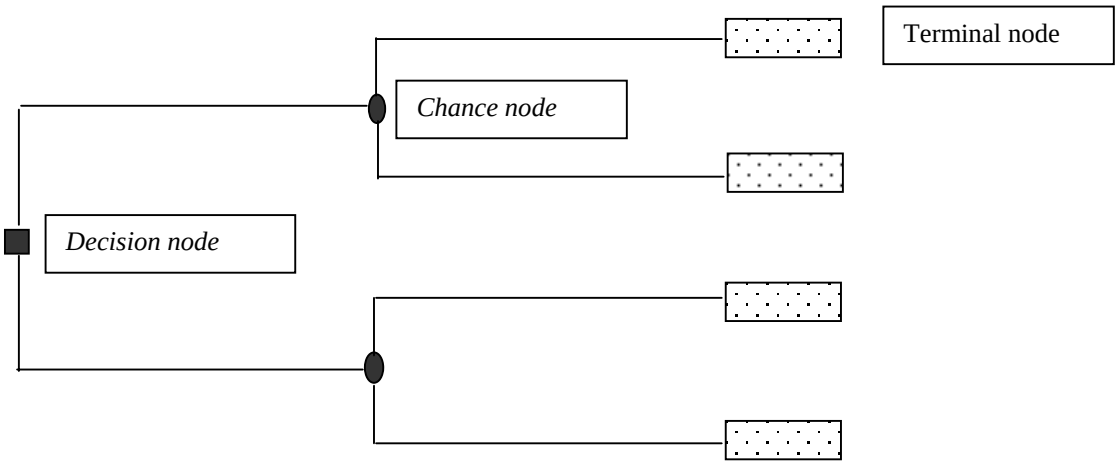

Fig. 4. Decision tree

Decision nodes **(boxes)** options available to the decision-maker: choice between 2 screening policies and/or between a screening policy *versus* no screening	***Chance nodes*** **(circles)** chance-driven events which are not under the decision-maker's direct control, such as an individual's compliance with screening, positive or negative test results, individuals returning for their results and follow-up or lost to follow-up, treatment offered that is accepted or refused, etc	***Terminal nodes*** **(rectangles)** represent the consequences of each decision pathway or final state of health: cost per case detected, cost per case detected and treated, etc.

VII.4. Conditions of model use

The reliability of a model is validated internally and externally.

- Internal validation examines how relevant are the data and how well they are organised into a structure.
- External validity is based on the simulated dynamics being appropriate representations of those actually observed.

As there are a number of hypotheses underlying the building of models, it is important to carry out a sensitivity analysis to verify the robustness of the results of the assessment with regard to variations in the key variables of the model (24).

VIII. Producing the final report

The UK National Screening Committee (7) has proposed a model report format which has the advantage of listing all the questions which need to be answered when evaluating a screening programme. This example could be used as the basis for a screening programme assessment report format.

I. Summary

II. Introduction

Purpose of the proposed screening programme– "Why screen for this disease?"

Systematic review method
- What question did the review address?
- Which populations were included/excluded?
- Which bibliographic databases were searched?
- Review strategy (keywords / MeSH searched)
- Reference list follow-up, personal contacts.
- Search for Non-English language / ongoing / unpublished studies.
- What inclusion/exclusion criteria were used for the studies?
- How was the quality of the studies assessed?
- What was the overall quality of the studies?
- Were the results from the studies combined? How was this done?

III. The health problem

Natural history of the disease
(including pathways of disease transmission, early symptomatic stage, recognisable latent period, disease markers)

Epidemiology of the disease
- Incidence, prevalence and projected trends.
- Mortality, morbidity and burden of disease by age/sex.
- "Is this an important health problem in comparison to other diseases?"

Primary prevention of the disease
- What are the opportunities/interventions for the primary prevention of the disease?
- How effective are these primary prevention interventions and what is the quality of the evidence?

IV. Current policy and practice

What is the current French policy on screening for the disease?
Describe the current French service (if any).

IV. The screening test

Describe the main screening tests and what they involve. What alternative tests are under consideration?

For each test, what is the distribution of test values in the target population? What is a suitable cut-off point and has this cut-off point been defined and agreed? Is there agreement on what constitutes a normal/abnormal/borderline test result?

For each test, what is the sensitivity, specificity, likelihood ratio and what is the quality of the evidence?

What are the side-effects/harmful effects of each test?

What is the acceptability of each screening test and what is the quality of the evidence?

VI. The diagnostic process

Diagnostic procedures
- What is the sequence of events for those who are positive on testing? (Describe the diagnostic process for positive individuals and the choices available to those individuals). Is there an agreed policy on the sequence of events?
- What are the diagnostic procedures and what do they involve?
- What are the side-effects/harmful effects of each diagnostic procedure?
- What is the acceptability of each diagnostic procedure and what is the quality of the evidence?

Is there an agreed policy as to which individuals should be offered treatment? State the policy.

VII. The treatment

What are the interventions and what do they involve?

What is the effectiveness of each intervention and what is the quality of the evidence?

Is there evidence that the treatment of patients identified through screening leads to better outcomes than later treatment?

What are the side-effects/harmful effects of each intervention?

What is the acceptability of each intervention and what is the quality of the evidence?

Is the quality of interventions and patient outcomes consistently high in all healthcare providers or is there evidence of variation in quality of care/patient outcomes?

VIII. The screening programme

What is the target population to whom screening will be offered?
- What proportion of potential cases are in the target population?
- What will be the positive rate at first screening?
- How best can the population be identified and targeted?

Is there evidence that the benefit-risk ratio of screening is satisfactory?

What is the proposed screening interval? (frequency with which the test is to be repeated).
Describe the evidence on interval disease progression and the rationale behind the proposed screening interval.

What level of patient uptake is required? (based on available evidence).

Present a decision analysis diagram of the pathway through the screening programme (from test to diagnosis to treatment/recall?)

IX. Beneficial effects

What are the benefits of screening for the disease?

What is the benefit-risk ratio?

What is the relative risk for the screened population compared to the control population? (for all cause and disease-specific mortality/morbidity).

What is the absolute risk reduction? (for all cause and disease-specific mortality/morbidity)

How does the benefit as a result of screening compare to that achieved in other screening programmes?

X. Adverse effects

What is the harm caused by the screening programme? (Including consequences of false-positive, false-negative, borderline results).

The physical harm.

The psychological harm.

XI. Absolute considerations

For every 100 000 individuals screened

- How many cases will be missed? (under-detection)
- How many will be treated? How does this compared to the number who would actually develop significant disease in the control group who were not offered screening? (over-detection).
- How many harmful effects will there be from the intervention?
- How many of the treated individuals will actually be helped? (i.e. in what proportion of screen-detected cases is an outcome improved?).
- How many individuals will be classified as borderline cases and what will happen to them?

Numbers needed to screen

- How many people have to be screened in order to find one treatable case?
- How many people have to be screened in order for one person to benefit?
- How many people are made anxious for each treatable case found? (false-positives and untreatable true positives)
- How many people are made anxious for one person to benefit?
- How many people are physically harmed for each treatable case found?
- How many people are physically harmed for one person to benefit?
- How many people are made anxious per 1000 persons screened?
- How many people are physically harmed per 1000 persons screened?
- How broad are the confidence intervals around the estimated size of the beneficial effect, and what are, at each end of the confidence intervals:
    - the number needed to screen;
    - the number adversely affected.

XII. Economic considerations

The costs of the screening programme
State the anticipated cost of the following if the screening programme was set up for a standard population of 10 million:

- setup costs;
- staff training;
- call up procedure;
- counselling;
- tests (and repeat tests);
- diagnostic procedures;
- intervention and follow-up;
- total setup and annual revenue/capital costs in order to deliver the programme for a standard population of 10 million.

What is the cost of finding one treatable case?
What is the cost in order for one person to benefit?
What are the potential savings which might result from the screening programme?
What is the cost effectiveness of the screening programme (and on what evidence is this based?)
Cost-benefit/cost-utility analysis.
What is the cost per QALY gained as a result of screening? How willing are individuals to pay?
Sensitivity analysis of screening for the disease
What implications does the screening programme have for other services?

XIII. Staffing and facilities

What are the clinical staffing implications of the screening programme? What will be the staffing requirements in order to introduce the screening programme for a standard French population of 10 million? Are sufficient numbers of clinical staff currently available or will further recruitment/training the required?

What facilities will be required in order to introduce the screening programme for a standard French population of 10 million?

XIV. Alternative options

What are the alternative policy options to screening?

What are the other ways of managing this health problem? (e.g. improving the treatment, providing other services)

How does the level of benefit as a result of screening compare to the benefit which could be achieved by improving individual diagnosis or treatment alone?

XV. Quality management

Who should manage the screening programme?

Quality assurance
- How should quality assurance be managed and monitored?
- What quality assurance standards should be recommended?

Describe an outline of the proposed service (equipment, siting, training, information needs of patients).

What are the critical success factors for the successful implementation of the screening programme?

XVI. Research

What relevant research is currently in progress?

Identify key areas for further research

XVII. Conclusions

General conclusions

Conclusions on each of the criteria for appraising screening programmes

Annex 1 – Participants

In house working group

Health Technology Assessment Department
Dr Stéphane Beuzon; Dr Emmanuel Corbillon; Dr Roselyne Delaveyne; Dr Etienne Duséhu; Dr Marie-Claude Hittinger; Dr Caroline Latapy; Dr Nathalie Poutignat; Dr Bertrand Xerri (Head of Department)

Economic Evaluation Department
Dr Fabienne Midy; Anne-Isabelle Poullié; Nathalie Préaubert; Catherine Rumeau-Pichon (Head of Department)

Other departments
Emmanuelle Blondet; Information specialist Documentation Department
Professor Isabelle Durand-Zaleski, Head of Professional Practice Appraisal Department
Sylvie Laot-Cabon, Project planning unit

External working group

Dr Rosemary Ancelle-Park, epidemiologist, Saint-Maurice
Dr Ségolène Aymé, Director of research into genetics and public health, Paris
Dr Claudine Blum-Boisgard, *Médecin-conseil*[1] at national level, Saint-Denis
Dr Georges Borges Da Silva, *Médecin-conseil,* Toulon
Dr Omar Brixi, public health physician, Montreuil
Dr Yves Charpak, Adviser to the director of WHO Regional Office for Europe, epidemiologist , Copenhagen
Professor Guy Delande, Professor of health economics, Montpellier
Dr Francis Fagnani, Director of CEMKA, a consultancy specialising in epidemiology, public health and health economics, Bourg-la-Reine
Dr Christine Jestin, public health physician, Paris
Dr Sandrine Loubière, researcher in health economics, Marseille
Professor Hélène Sancho-Garnier, Professor of public health, Montpellier

Peer reviewers

Members of the ANAES Scientific Council
Professor Marie-Odile Carrère, Professor of economics, Lyon
Dr Michel Cucherat, Pharmacologist, Biostatistics Department and UCBL1 biostatistics laboratory, Lyon
Professor Maryse Gadreau, Professor of economics, Dijon
Dr Pierre Haehnel, radiologist, Strasbourg
Professor Paul Landais, Medical information and biostatistics department, Necker-Enfants Malades Hospital, Paris
Professor Alain Vergnenègre, chest physician, Limoges

Others
Tiiu Ojasoo Ph.D., Scientific editor, ANAES, Saint-Denis
Professor Roger Salamon, University hospital medical information department, Bordeaux

Annex 2 - Glossary of economic terms

SENSITIVITY ANALYSIS – Analysis used to test the robustness of a model. Within defined limits, the analysis shows how outcomes are altered if the main parameters on which the model is built are changed.

DIRECT COSTS – From a purely logical point of view, costs classed as direct costs are those directly attributable to the disease, its treatment and/or primary or secondary preventive measures against it. Costs may be medical resources (hospitalisation, outpatient care, drugs etc.) or non-medical costs (non-ambulance transport, childcare costs, cost of adapting the home, time of the natural carers, etc.)

INDIRECT COSTS – Sum of the potential work the patient would have done if they had not been ill, expressed as a monetary unit.

MARGINAL COST – The extra cost needed to produce one more service/product (for example the cost of the last day spent in hospital, or the cost of an additional patient screened).

AVERAGE COST – Cost per unit produced: relationship between the total cost and the number of units produced.

OPPORTUNITY COST – The true concept of cost in economics, the opportunity cost of a programme represents the health results that would have been obtained by another programme if the resources used had been allocated to the latter rather than to the former.

EFFICIENCY – Relationship between the result obtained and the resources used. A policy is said to be efficient when it makes it possible to obtain the maximum result for a given cost, or when it makes it possible to minimise the costs for a given result.

COST-BENEFIT ANALYSIS – Economic evaluation method relating the monetary costs of a health project and its results expressed in monetary units (certain aspects of interventions are difficult to express in monetary terms, e.g. pain). As the costs and benefits are expressed in monetary terms, the usual economic criteria for investment may be applied to a cost-benefit analysis.

COST-EFFICACY ANALYSIS – Cost-effectiveness analyses are used to establish an efficient policy. Effectiveness is measured by a clinical outcome indicator or an object state of health indicator (death, life-years, etc.). They also give decision-makers information about additional effectiveness obtained through additional cost.

COST-UTILITY ANALYSIS – Particular form of cost-effectiveness analysis. The medical results are expressed in terms of equivalent quality-adjusted life-years (QALY) or by other indicators of utility.

COST-MINIMISATION ANALYSIS – Cost minimisation analyses are used when the interventions being compared differ only in terms of the cost involved. When two policies are equally effective with regard to treatment and have the same consequences (medical and social) for the patient, but differ in cost, the cheapest intervention is sought.

COST-RESULT ANALYSIS – Group of methods used to establish a relationship between the cost of different interventions and their results (cost-minimisation analysis, cost-effectiveness analysis, cost-utility analysis, cost-benefit analysis).

CLINICAL AND ECONOMIC EVALUATION – A clinical and economic evaluation tries to establish a relationship between the cost of different interventions and their results. It is the tool used to evaluate the efficiency (i.e. the relationship between cost and performance) of medical policies. From a macroeconomic

perspective, it establishes the benefit of a medical activity by establishing the relationship between the cost for the community and the advantages for the population concerned by this activity.

DISCOUNT RATE – Used to convert future costs and benefits into current values. A reflection of the values that society places in the future.

EXTERNAL VALIDITY – Assurance that the results of a study can be generalised to another similar population. In epidemiology this term is replaced by the term “generalizability”.

INTERNAL VALIDITY – Assurance that a study or tool can measure what it claims to measure. Term used in sociology as equivalent to the term validity in epidemiology.

Sources (for original French glossary):
http://lexeco.free.fr/indexf.htm, http://www.bdsp.tm.fr/Glossaire/Default.asp, Drummond et al. (26), Grignon & Midy (42), Béresniak & Duru (43), Collège des économistes de la santé (24), Durieux (44).

Annex 3. Method used to produce this report

This guide was based on a literature review and the opinion of the members of two working groups (see Annex 1). The in-house working group consisted of ANAES project managers and heads of department; the external working group consisted of 11 members. The final report was submitted to 8 peer reviewers for comments.

The Medline database was searched (National Library of Medicine, United States) as follows:

Study type/subject	MeSH descriptors or terms in titles or summaries	**Search period**
Appraisal of screening programmes		Unlimited
STEP 1 AND STEP 2	*Screening* OR *Mass screening* *Program Evaluation* OR *Public Health* OR *Health Priorities* OR *Health Planning* OR *Health Planning Guidelines* OR *Health Services Research*	
Methodological aspects of economic evaluation		
STEP 1 AND STEP 3	*(Method* OR *Program Evaluation* OR *Public Health* OR *Health Priorities* OR *Health Planning* OR *Health Planning Guidelines* OR *Health Services Research)* AND *(Cost allocation* OR *Cost-benefit analysis* OR *Cost control* OR *Cost of illness* OR *Cost savings* OR *Costs and cost analysis* OR *Cost effectiveness* OR *Economic value of life* OR *Health care cost* OR *Health economics* OR *Economic aspect* OR *Hospital cost* OR *Hospital charge* OR *Financial management, hospital* OR *Hospital billing* OR *Hospital finance* OR *Hospital running cost* OR *Pharmacoeconomics* OR *Cost(s)* OR *Economic(s))*	
Ethical aspects of screening		Unlimited
STEP 1 AND STEP 4	*Ethics*	
	Number of articles found 440 **Number of articles examined 108**	

Websites of the following agencies were consulted:

- *Haut Comité de la santé publique*: hcsp.ensp.fr
- *Société française de santé publique*: www.sfsp-publichealth.org
- INAHTA (International Network of Health Agencies for Health Technology Assessment Department): www.inahta.org
- UK National Screening Committee: www.nsc.nhs.uk
- *Groupe canadien pour l'examen médical périodique*: www.ctfphc.org
- NHS Centre for Reviews & Dissemination: www.york.ac.uk
- National Institute for Clinical Excellence: www.nice.org.uk
- US Centers for Disease Control: www.cdc.gov
- US Agency for Healthcare Research & Quality: www.ahrq.gov
- New Zealand Screening Unit: www.healthywomen.org.nz

Annex 4. WHO criteria for a screening programme

The following list of criteria is taken from “Principles and practice of screening for disease”, WHO, 1970 (8).

1. The condition sought should be an important health problem for the individual and community.

2. There should be an accepted treatment or useful intervention for patients with the disease.

3. Facilities for diagnosis and treatment should be available.

4. There should be a latent or early symptomatic stage.

5. There should be a suitable and acceptable screening test or examination.

6. The test used should be acceptable to the population.

7. The natural history of the disease should be adequately understood, including development from latent to declared disease.

8. There should be an agreed policy on which individuals should be treated.

9. The cost 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.

Annex 5. STARD checklist for studies of diagnostic accuracy (17)

Title/abstract/ keywords	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').
Introduction	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.
Methods:		
participants	3	The study population: The inclusion and exclusion criteria, setting and locations where the data were collected.
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?
Test methods	7	The reference standard and its rationale.
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.
	9	Definition of and rationale for the units, cutoffs and/or categories of the results of the index tests and the reference standard.
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).
	13	Methods for calculating test reproducibility, if done.
Results:		Report
participants	14	When study was done, including beginning and ending dates of recruitment.
	16	The number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended).
Test results	17	Time interval from the index tests to the reference standard, and any treatment administered between.
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.
	20	Any adverse events from performing the index tests or the reference standard.
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).
	22	How indeterminate results, missing responses and outliers of the index tests were handled.
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.
	24	Estimates of test reproducibility, if done.
Discussion	25	Discuss the clinical applicability of the study findings.

Annex 6. 10-point checklist (Drummond et al. (26))

1. Was a specific question posed?
1.1. Did the study take account of both the costs and the results of the programme(s)?
1.2. Did the study compare different options?
1.3 Was a specific viewpoint adopted and was the study positioned in a particular context in relation to decision-making?

2. Were competing alternatives described in detail? (i.e. can you say who? did what? to whom? where? and how often?)
2.1. Were important alternatives omitted?
2.2. Was a "do nothing" alternative included (should it have been?)?

3. Has the effectiveness of the programmes been established?
3.1. Was the evaluation based on a randomised controlled clinical trial? If so, did the trial protocol reflect what would normally happen in current practice?
3.2. Was effectiveness established from a review of clinical trials?
3.3. Were observational data or hypotheses used to establish effectiveness? If so, in what way were the results biased?

4. Were the most important costs and consequences for each alternative identified?
4.1. Was the research field sufficiently broad for the question posed?
4.2. Were the different relevant points of view examined (e.g. the point of view of society; point of view of the patient and of the third-party payer; other points of view could be relevant in a given analysis)
4.3. Were capital and running costs included?

5. Were costs and consequences measured correctly, in appropriate physical units? (e.g. number of hours of nursing care, number of consultations, working days lost, life-years gained)
5.1. Have all the items identified been measured? When an item has been discarded, could it be considered to be negligible?
5.2. Were there any special circumstances (for example, shared use of resources) which would make the calculation difficult? Have they been properly taken into account?

6. Were the costs and outcomes adjusted for time?
6.1. Were future costs and outcomes discounted?
6.2. Was the choice of discount rate justified?
6.3. How were market prices estimated when they were missing (e.g. in the case of work by volunteers) or when they did not reflect real values (e.g. in the case of a subsidised healthcare organisations)?
6.4. Was the evaluation of consequences appropriate for the question posed (i.e. were adequate analyses done, such as cost-effectiveness, cost-utility, cost-benefit)?

7. Were the costs and outcomes adjusted for time?
7.1. Were future costs and outcomes discounted?
7.2. Was the choice of discount rate justified?

8. Was a differential analysis carried out of costs and outcomes for competing alternatives?
Were the additional costs generated by one alternative rather than another compared with its additional effects, benefits or utilities?

9. Was uncertainty taken into account in the estimate of costs and outcomes?
9.1. If the data on costs and outcomes were stochastic, were appropriate statistical analyses performed?
9.2. If a sensitivity analysis was carried out, were the value ranges (for key variables) justified?
9.3. What is the sensitivity of the study results to changes in variables (for values used in the sensitivity analysis, or within the confidence interval of the ratio of costs to outcomes)?

10. Did the presentation and discussion of the study results cover all users' concerns?
10.1. Were the study's conclusions based on a particular overall indicator (e.g. a cost-effectiveness ratio)? In this event, was it interpreted correctly?
10.2. Were the results compared with those from other studies on the same subject? In this case, were possible differences in design taken into account?
10.3. Did the study address the issues of generalising the results, for different contexts or different groups of patients?
10.4. Did the study mention or take account of other important factors relating to the decision in question (e.g. distribution of costs and outcomes, or ethical issues)?
10.5. Did the study address the problems posed by the implementation of the chosen programme, taking account of financial or other constraints, and was the question raised whether resources that may have been freed up could be reallocated to other programmes of value?

References

1. Agence nationale d'accréditation et d'évaluation en santé. Évaluation d'une action de santé publique : recommandations. Paris: ANAES; 1995.

2. Durand-Zaleski I, Bastuji-Garin S. Évaluation des procédures de diagnostic ou de dépistage. Validité d'un test, sensibilité, spécificité, valeurs prédictives, définitions et indications d'un dépistage de masse. Rev Prat 2000;50:1155-8.

3. Council of Europe, Committee of Ministers, Recommendation No. R (94) 11 on Screening as a Tool of Preventive Medicine Brussels, Council of Europe; 1994.

4. Charpak Y. Quelques suggestions préalables à la réalisation d'études d'évaluation : pour une "commande" optimale. Santé Publique 1998;10(2):225-30.

5. Claveranne JP. L'évaluation : nature et formalisation d'un concept mou. Rev Cahier Lyonnais Recherche Gestion 1994;15:232-46.

6. Conseil scientifique de l'évaluation. Petit guide de l'évaluation. Paris: CSE; 1990.

7. The national screening committee handbook of population screening programmes. London: National Screening Committee; 1998.

8. Wilson JMG, Jungner G. Principes et pratique du dépistage des maladies. Genève: Organisation mondiale de la santé; 1970.

9. Cochrane AL, Holland WW. Validation of screening procedures. Br Med Bull 1969;27(1):3-8.

10. Sackett DL, Holland WW. Controversy in the detection of disease. Lancet 1975;2:357-9.

11. Grenier B. Évaluation de la décision médicale : introduction à l'analyse médico-économique. Paris: Masson; 1996.

12. Beaucage C, Viger BY. Épidémiologie appliquée. Une initiation à la lecture critique de la littérature en sciences de la santé. Montréal: Gaëtan Morin; 1996.

13. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Bases Medicine Working group. JAMA 1994;271(9):703-7.

14. National Health Medical Research Council. Child health screening and surveillance: a critical review of the evidence. Canberra: NHMRC; 2002.

15. Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K. False-negative results in screening programmes: systematic review of impact and implications. Health Technol Assess 2000;4(5):1-70.

16. Salmi LR, Collet JP. Lecture critique des articles médicaux. II. Juger de l'intérêt d'un test diagnostique. Rev Prat 1991;41(25):2734-43.

17. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM *et al.* Towards complete and accurate reporting of studies of diagnosis accuracy: the STARD initiative. BMJ 2003;326:41-4.

18. Touzet S, Chapuis F, Colin C. Aspects méthodologiques de l'évaluation du dépistage : à propos du dépistage de l'hépatite virale C. Gastroentérol Clin Biol 2000;24:631-6.

19. Dubois G. Principes de l'évaluation d'un test de dépistage. J Alcoologie 1989;1:7-21.

20. Barratt A, Irwig L, Glasziou P, Cumming RG, Raffle A, Hicks N *et al.* Users' guide to the medical literature XVII. How to use guidelines and recommendations about screening. JAMA 1999;281(21):2029-34.

21. Campbell MK, Elbourne DR, Altman DG. CONSORT statement: extension to cluster randomised trials. BMJ 2004; 328:702-8.

22. Giraudeau B. L'essai clinique randomisé par grappes. Méd Sci 2004;2:263-6.

23. Earle C, Hébert PC. A reader's guide to the evaluation of screening studies. Postgrad Med J 1996;72:77-83.

24. Collège des économistes de la santé. Guide méthodologique pour l'évaluation économique des stratégies de santé. Recommandations méthodologiques. Paris: CES; 2003.

25. Moatti JP. Contribution de l'analyse économique au débat sur le dépistage des cancers : des faux alibis aux vrais dilemmes. In: Sancho-Garnier H, Béraud C, Doré JF, Pierret J, Schaffer P, ed. Dépistage des cancers. De la médecine à la santé publique. Paris: Éditions Inserm 1997. p. 93-107.

26. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. Méthodes d'évaluation économique des programmes de santé. Paris: Economica; 1998.

27. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.

28. Canadian Coordinating Office for Health Health Technology Assessment Department. Lignes directrices pour l'évaluation économique des produits pharmaceutiques : Canada. Ottawa: CCOHTA, 2e éd.; 1998.

29. Segnan N, Armaroli P, Sancho-Garnier H. Dépistage. In press.

30. Moatti JP, Loubière S, Rotily M. L'analyse économique face au principe de la garantie de sécurité en transfusion sanguine. Transf Clin Biol 2000;7:228-35.

31. Mamelle N, Lacour A, Anes A, Bazin B, Chaperon J, Duru G *et al.* Les expérience de dépistage de masse du cancer du sein par mammographie en France. Un protocole commun d'évaluation. Rev Epidémiol Santé Publique 1994;42(1):34-49.

32. Brown J, Buxton M. The economic perspective. Br Med Bull 1998;54(4):993-1009.

33. Bilger P. Analyse économique du dépistage de masse du cancer du sein [thesis]. Strasbourg : université Louis-Pasteur; 1997.

34. Kobelt G. L'économie de la santé : une introduction à l'évaluation économique. London: Office of Health Economics; 1997.

35. Provenzale A. A reader's guide to economic analysis in the GI literature. 1996.

36. Castiel D. Efficacité des procédures d'allocation des ressources : quand l'équité vient au secours de l'économie. Cah Gratice 2000;15:147-67.

37. Sancho-Garnier H, Lancry PJ, Fagnani F. Le dépistage des cancers : pour que la fin justifie les moyens. Méd Sci 1992;8:10-5.

38. Agence nationale d'accréditation et d'évaluation en santé. Élaboration d'un document d'information à l'intention des patients ou usagers du système de santé. Guide méthodologique. Paris: ANAES; 2004.

39. Ligue nationale contre le cancer. Comment réussir une campagne de dépistage de masse des cancers du sein. Paris: LNCC; 1994.

40. Markham IS. Ethical and legal issues. Br Med Bull 1998;54(4):1011-21.

41. Moto L. Santé et multidisciplinarité : choix et décisions. Paris: Hermès; 1995.

42. Grignon M, Midy F. La notion de coût en économie de la santé. Document de travail. Paris: Credes; 2001.

43. Béresniak A, Duru G. Économie de la santé. Paris: Masson; 2001.

44. Durieux P. Guide des principaux termes dans le domaine de la santé. Évaluation, qualité, sécurité. Paris: Flammarion; 1997.