

Glossary

Background incidence rate:	The breast cancer incidence rate expected in the absence of screening
Invasive breast cancer detection rate:	The number of histologically proven malignant lesions of the breast (invasive) detected at screening per 1000 women
Total breast cancer detection rate:	The number of histologically proven malignant lesions of the breast: in-situ (ductal only, not lobular) and invasive detected at screening per 1000 women
Breast cancer incidence rate:	The rate at which new cases of breast cancer occurs in a population. The numerator is the number of newly diagnosed cases of breast cancer that occurs in a defined period. The denominator is the population at risk of a diagnosis of breast cancer during this defined period, sometimes expressed in person-time.
Breast cancer mortality rate:	The rate at which deaths from breast cancer occur in a population. The numerator is the number of breast cancer deaths that occurs in a defined time period. The denominator is the population at risk of dying from breast cancer during this defined period, sometimes expressed as person-time.
Breast cancer register:	Recording of information on all new cases of and deaths from breast cancer occurring in a defined population
Delay time:	The time between when a cancer could be detected by screening and when it is actually detected
Efficacy:	The reduction in breast cancer mortality in randomized trials, under ideal conditions
Effectiveness:	The reduction in breast cancer mortality in screening practice, under real conditions
Eligible population:	The adjusted target population, i.e. the target population minus those women who are excluded according to screening policy on the basis of eligibility criteria other than age, sex and geographical location
Further assessment:	Additional diagnostic steps (either non-invasive or invasive) performed to clarify the nature of an abnormality detected at screening, either at the time of screening or on recall
Interval cancer:	A primary breast cancer diagnosed in a woman who had a result in a screening test, with or without further assessment, that was negative for malignancy, either: <ul style="list-style-type: none">• before the next invitation to screening was due or• within a period equal to a screening interval for a woman who has reached the upper age limit for screening

Interval cancer rate:	The number of interval cancers diagnosed within a defined period since the last negative result in a screening examination per 1000 women with negative results
Lead time:	Period between when a cancer is found by screening and when it would be detected from clinical signs and symptoms (not directly observable)
Length bias:	The bias towards detection of cancers with longer sojourn times and therefore a better prognosis by screening
Open biopsy:	Surgical removal of (part of) a breast lesion
Organized screening:	Screening programmes organized at national or regional level, with an explicit policy, a team responsible for organization and for health care and a structure for quality assurance
Opportunistic screening:	Screening outside an organized or population-based screening programme, as a result of e.g. a recommendation made during a routine medical consultation, consultation for an unrelated condition, on the basis of a possibly increased risk for developing breast cancer (family history or other known risk factor) or by self-referral
Overdiagnosis:	Detection of breast cancers that might never have progressed to become symptomatic during a woman's life
Participation rate:	Number of women who have a screening test as a proportion of all women who are invited to attend for screening
Population access:	Proportion of the national population of eligible women who have access to a screening programme
Positive predictive value:	Proportion of all positive results at screening that lead to a diagnosis of cancer
Recall:	Physical recall of women to the screening unit, as a consequence of the screening examination, for: <ul style="list-style-type: none">• a repeat mammogram because of technical inadequacy of the screening mammogram (technical recall); or• clarification of a perceived abnormality detected at screening, by performance of an additional procedure (recall for further assessment).
Recall rate:	The number of women recalled for further assessment as a proportion of all women who were screened
Refined mortality:	Mortality rate among women, excluding those in whom breast cancer was diagnosed before screening began
Screening interval:	Fixed interval between routine screenings decided upon in each programme, depending on screening policy
Screening policy:	Specific policy of a screening programme which dictates the targeted age and sex group, the geographical area, the screening interval (usually 2 or 3 years), etc.

Screening test:	Test applied to all women in a programme, consisting of a single or two-view mammogram with or without clinical examination
Sensitivity:	The proportion of truly diseased persons in the screened population who are identified as diseased by the screening test. The more general expression for 'sensitivity of the screening programme' refers to the ratio of true positives (breast cancers correctly identified at the screening examination) / true positives + false negatives (breast cancers not identified at the screening examination, detected as interval cases).
Sojourn time:	Detectable preclinical phase; time between that at which a tumour could be found by screening and that at which it would appear symptomatically (not directly observable)
Specificity:	Proportion of truly non-diseased persons in the screened population who are identified as non-diseased by the screening test (i.e. true negatives / true negatives + false positives)
Target population:	The age-eligible population for screening, e.g. all women offered screening according to the policy

Working Procedures

Prevention of cancer is one of the key objectives of IARC. Secondary prevention by early diagnosis and screening in non-symptomatic individuals is a fundamental component of any cancer control programme. The aim of secondary prevention is to reduce mortality and suffering from the disease. When screening is planned as part of a cancer control programme, only strategies proved to be effective should be proposed to the general population. Screening usually requires repeated interactions between 'healthy' individuals and health care providers, which can be inconvenient and costly. Furthermore, screening requires an ongoing commitment between the public and health care providers.

Scope

Cochrane (1972) first discussed the concepts of efficacy and effectiveness in the context of health interventions. Efficacy was later defined by Last (1995) as "the extent to which a specific intervention, procedure or service produces a beneficial result under ideal circumstances". In contrast, the related term "effectiveness" is defined by the same author as "... a measure of the extent to which a specific intervention, procedure, regimen or service, when deployed in the field in routine circumstances, does what it is intended to do for a specific population." The distinction between efficacy as measured in experimental studies and the effectiveness of a mass population intervention is a crucial one for public health decision-making. In particular, the fact that the effectiveness of a screening procedure may be differ-

ent in different populations is often overlooked. A mass programme of screening must satisfy certain minimal requirements (e.g. acceptability, availability of relevant personnel, facilities for screening and access to pertinent health services) if it is to achieve the results that have been documented in randomized trials. The acceptance and use of screening services may vary from one population to another, implying that a given screening procedure is not universally effective. Even when a screening procedure is effective as a mass intervention, other outcomes such as harms and costs and the potential for other interventions to achieve equivalent benefits must be considered.

Efficacy is a necessary but not a sufficient basis for recommending screening. The efficacy of a screening procedure can be inferred if effectiveness can be proven. Screening has sometimes been implemented by a given procedure on the assumption that 'earlier is better', even when no evidence of efficacy was available. If such interventions result in a significant reduction in mortality that cannot otherwise be explained (by reduced incidence, due perhaps to primary prevention or better treatment), it can be inferred that the procedure is effective. However, uncontrolled interventions in which individuals are exposed to unknown risks and benefits should be avoided.

Objectives

The objectives of the Working Group are:

- (1) to evaluate the strength of the evidence for the efficacy of a screening procedure;
- (2) to assess the effectiveness of defined screening interventions in defined populations;
- (3) to assess the balance of benefit and harm in target populations; and
- (4) to formulate recommendations for further research and for public health action.

The conclusions of the Working Group are published as a volume in the series of the IARC Handbooks of Cancer Prevention.

Working groups

An international working group of experts is convened by the IARC. The tasks of the group are:

- (1) to ascertain that all appropriate data have been retrieved;
- (2) to select the data relevant for evaluation on the basis of scientific merit;
- (3) to prepare accurate reviews of data to allow the reader to follow the reasoning of the working group;
- (4) to evaluate the efficacy and effectiveness of the screening procedure;
- (5) to summarize the potential adverse consequences of screening;
- (6) to prepare recommendations for research and for public health action; and
- (7) to prepare an overall evaluation of the screening procedure at the population level.

Approximately 13 months before a working group meets, the topics of the

Handbook are announced, and prospective participants are selected by IARC staff in consultation with other experts. Subsequently, relevant data are collected by the IARC from all available sources of published information. Working Group participants who contributed to the considerations and evaluations within a particular handbook are listed, with their addresses, at the beginning of each publication. Each participant serves as an independent scientist and not as a representative of any organization, government or industry. They are expected to put aside any stake they may have in a particular outcome and to evaluate the evidence objectively and with scientific rigour. Scientists nominated by national and international agencies, industrial associations and consumer and/or environmental organizations may be invited as observers. IARC staff members involved in the preparation of the handbook are listed.

About eight months before the meeting, the material collected is sent to meeting participants who are asked to prepare sections for the first drafts of the handbook. These drafts are then compiled by IARC staff and sent, before the meeting, to all participants of the working group for review.

Data for handbooks

The handbooks do not necessarily cite all of the literature on the agent or strategy being evaluated. Only those data considered by the working group to be relevant to making the evaluation are included. Meeting abstracts and other reports that do not provide sufficient detail upon which to base an assessment of their quality are generally not considered.

With regard to reports of basic scientific research, epidemiological studies and clinical trials, only those that have been published or accepted for publication in the openly available scientific literature are reviewed by the working group. In certain instances, government

agency reports that have undergone peer review and are widely available are considered. Exceptions may be made ad hoc to include unpublished reports that are in their final form and publicly available, if their inclusion is considered pertinent to making a final evaluation.

The available studies are summarized by the working group. In general, numerical findings are indicated as they appear in the original report; units are converted when necessary for easier comparison. The working group may conduct additional analyses of the published data and use them in their assessment of the evidence. These analyses are described in the handbook. Important aspects of a study, directly impinging on its interpretation, are brought to the attention of the reader.

Evaluation of screening

The framework of a handbook on screening includes the following nine chapters:

Chapter 1. Disease characteristics, global burden and rationale for screening

Descriptive epidemiology

The purpose of this section is to document the importance of the disease in the context of the general health status of different populations. The worldwide burden of the cancer is described (mortality, incidence, prevalence and survival rates) and integrated with measures of the occurrence of cancers at other sites, of mortality from all causes and life expectancy. Expected trends in the absence of screening are a relevant component of this section.

Natural history of the disease as relevant to screening

In this section, the natural history of the disease of interest and the relevance and potential of screening for early detection and for reducing mortality are described. Evolving concepts and

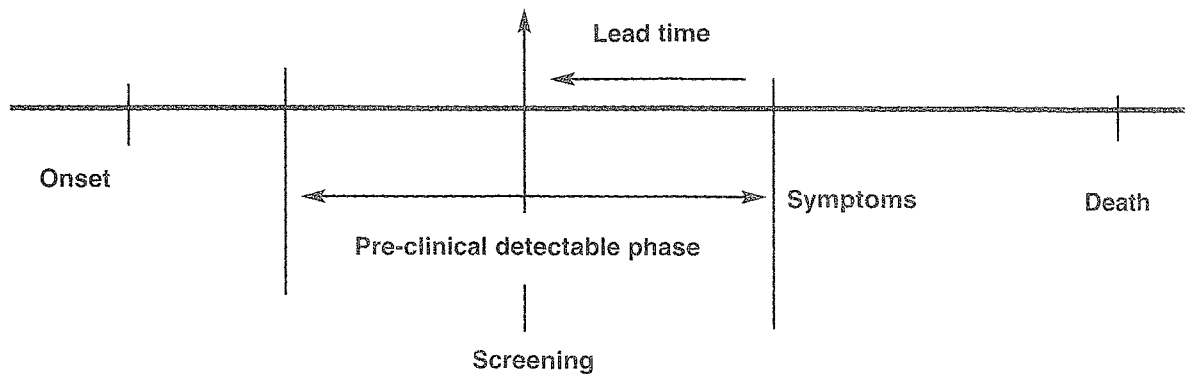
principles pertinent to screening are also discussed.

There is now a wealth of evidence (both direct and indirect) to support the principle that screening and detection of certain cancers in appropriate target populations are associated with a lower probability of dying from the disease. The scheme (on the next page) illustrates the temporal framework commonly subscribed to in modern screening models

It should be noted that early diagnosis, due to greater awareness and improved access to appropriate medical services, has resulted in many countries in a reduction in diagnostic delay, probably reducing mortality. As a consequence, symptomatic cancers are frequently diagnosed and treated early after the onset of symptoms in many developed nations. In such instances, screening for the disease will improve outcomes (such as a reduction in mortality) only if treatment of the disease at an even earlier phase in its development provides additional benefit. The rapid evolution of molecular or genetic markers of pre-malignant conditions or individuals at 'high risk' has modified the concepts of 'disease onset' and 'lead time'. Hence, the model outlined above may require adaptation or development to allow for detection of pre-clinical conditions of undetermined significance (including serological and molecular markers and genetic predisposition), if they are relevant for screening for the cancer in question.

Chapter 2. Screening tests

It is important to distinguish between screening tests and screening procedures, i.e. the test itself and the way in which it is administered. The two merit separate, detailed evaluation. Each of the screening tests to be considered is described. The ability of each test to detect cancer and to distinguish cancer from non-cancer conditions will be assessed as:



- the validity of the test, expressed as its sensitivity and specificity under various conditions;
- all known or potential side-effects; and
- the cost of the test when implemented in mass screening programmes.

Chapter 3. Delivery and uptake of screening

Information on how screening is delivered in different countries is reviewed in this section, with emphasis on the following aspects:

- infrastructure for diagnosis and treatment: the nature of standard diagnostic procedures and treatment regimens and their availability to the target population;
- extent of population coverage and participation rates;
- equity, as defined by the extent to which access to the procedure (including diagnostic investigation and treatment) is ensured for all eligible individuals, irrespective of any personal characteristics;
- informed decision and informed consent: the extent to which individual values are respected when information on potential benefit and harms is conveyed; and

- behavioural and demographic considerations that affect participation in screening.

Chapter 4. Efficacy of screening tests

In this section, evidence from experimental and observational studies is reviewed, and aspects of study design and analysis are critically discussed.

The handbooks are not intended to summarize all published studies. The working group considers the following aspects:

- (1) the relevance of the study;
- (2) the appropriateness of the design and analysis to the question being asked;
- (3) the adequacy and completeness of the presentation of the data; and
- (4) the degree to which chance, bias and confounding may have affected the results.

Studies that are judged to be inadequate or irrelevant to the evaluation are generally omitted. They may be mentioned briefly (i) when the information is considered to be a useful supplement to that in other reports, (ii) if they provide the only data available or (iii), in exceptional cases, if they have been widely perceived as being pertinent but are deemed otherwise by the Working

Group. Their inclusion does not imply acceptance of the adequacy of the study design nor of the analysis and interpretation of the results, and their limitations are outlined.

The appropriate outcome(s) (mortality or incidence) of a given procedure, e.g. the detectable phase(s) of the natural history of the disease, are also defined. Aspects that are particularly important in evaluating experimental studies are: the selection of participants, the nature and adequacy of the randomization procedure, evidence that randomization achieved an adequate balance between the groups, the exclusion criteria used before and after randomization, compliance with the intervention in the screened group and 'contamination' with the intervention in the control group. Other considerations are the means by which the endpoint was determined and validated (either by screening or by other means of detection of the disease), the length and completeness of follow-up of the groups and the adequacy of the analysis.

Whenever possible, similar criteria should be used to evaluate non-experimental comparative studies.

In the Working Group's analysis of the efficacy of the screening procedure, a meta-analysis may be used, when applicable.

In evaluating case-control and cohort studies, particular attention is paid to the definition of cases, controls and exposure and, for cohort studies, the length and completeness of follow-up. Potential bias, especially selection bias, is carefully examined in all observational studies.

Chapter 5. Effectiveness of population-based screening

The impact of the screening procedure when implemented in defined populations is examined in this section. Indicators used to monitor effectiveness, such as positive and negative predictive values, detection rate, rates of interval cancers and the number of tests performed, are reported. Time trends before and after implementation of screening as well as geographical comparisons of the occurrence of the disease and death from the disease in populations exposed and not exposed to screening are reviewed and interpreted. In doing this, the Working Group takes into account differences in screening procedures (e.g. frequency and the age of the target population) and of participation rates.

An integral component of this section is an evaluation of the benefits and harms of the screening procedure to the population. Reductions in mortality and/or incidence of invasive disease are fundamental measures of benefit. A reduction in the cumulative prevalence of advanced disease may be a useful predictor of a reduction in mortality from the disease. An additional benefit may be that more cases can be treated by less aggressive, less invasive procedures, thus improving the quality of life.

The spectrum of health care is dynamic, and a screening procedure should not be viewed in isolation. Greater awareness of the disease, brought about by publicity about screening that may result in early diagnosis, could be regarded as another

benefit of a screening programme. This section should also consider the possibility that there might have been a change in treatment of the cancer, which even in the absence of screening would have resulted in a substantial decrease in mortality. As far as possible, an evaluation should be made of the extent to which improved treatment has been responsible for any changes seen in mortality from the specific disease.

Estimates of the rates of false-positive and false-negative findings in screened individuals and their consequences (false sense of security with false-negatives and false alarm with false-positives) are an integral part of this section. The rates of short- and long-term side-effects and the possibility of unnecessary treatment of borderline or indolent cases detected at screening are discussed. Management procedures for lesions detected at screening are reviewed. Psychological factors, such as anxiety induced by undergoing the test procedure, are also considered.

Chapter 6. Cost-effectiveness of population-based screening

In this section, the cost-effectiveness of various modalities of test administration in various settings is considered. The discussion takes into account the costs per case detected and per death prevented.

Chapter 7. Summary of data

In this section, the relevant data are summarized. Inadequate studies identified in the preceding text are generally not included.

Chapter 8. Evaluation

Evaluation of the efficacy of the screening procedure

An evaluation of the degree of evidence for the efficacy of a screening procedure is formulated according to the following definitions:

Sufficient evidence of the efficacy of cancer-preventive activity will apply when screening interventions by a defined procedure are consistently associated with a reduction in mortality from the cancer and/or a reduction in the incidence of invasive cancer, and chance and bias can be ruled out with reasonable confidence.

Limited evidence of the efficacy of cancer-preventive activity will apply when screening interventions by a defined procedure are associated with a reduction in mortality from the cancer and/or a reduction in the incidence of invasive cancer or a reduction in the incidence of clinically advanced cancer, but bias or confounding cannot be ruled out with reasonable confidence as alternative explanations for these associations.

Inadequate evidence of the efficacy of cancer-preventive activity will apply when data are lacking or when the available information is insufficient or too heterogeneous to allow an evaluation.

Sufficient evidence that the screening procedure is not efficacious in cancer prevention will apply when any of the following cases hold:

- the test does not result in earlier diagnosis than with standard tests already in use;
- the survival of cases detected at screening is no better than that of cases diagnosed routinely;
- the screening interventions are consistently associated with no reduction in mortality from the cancer, and bias can be ruled out with reasonable confidence.

In case of limited or inadequate evidence, the Working Group should highlight those aspects of the procedure for which information is lacking and which led to the uncertainty in evaluation. This will provide indications of research priorities.

Overall evaluation

Finally, the body of evidence is considered as a whole, and summary statements are made about the cancer-preventive effects of the screening intervention in humans and other beneficial or adverse effects, as appropriate. The overall evaluation is usually in the form of a narrative. The data on the effectiveness of the screening intervention are summarized, including the factors that determine its success and failure under

routine conditions. Finally, the balance between expected benefit and harm is described.

Chapter 9. Recommendations

After its review of the data and its deliberations, the working group formulates recommendations, where applicable, for:

- further research and
- public health action.

References

- Cochrane, A.L. (1972) *Effectiveness and Efficiency: Random Reflections on Health Services*, Oxford: Nuffield Provincial Hospitals Trust
- Last, J.M. (1995) *A Dictionary of Epidemiology*, Oxford: Oxford University Press

Sources of figures

Figure 1	Ferlay <i>et al.</i> (2001)	Figure 14	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 27	Working Group
Figure 2	Ferlay <i>et al.</i> (2001)	Figure 15	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK,	Figure 28	Blackman <i>et al.</i> (1999)
Figure 3	Descriptive Epidemiology Unit (International Agency for Research on Cancer) (DEP/IARC)	Figure 16	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 29	Nyström <i>et al.</i> (2002)
Figure 4	DEP (IARC)	Figure 17	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 30	Ford <i>et al.</i> (1998)
Figure 5	DEP (IARC)	Figure 18	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 31	WHO (1999)
Figure 6	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 19	Working Group	Figure 32	WHO (1999)
Figure 7	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 20	Working Group	Figure 33	WHO WHO (1999)
Figure 8	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 21	Working Group	Figure 34	WHO (1999)
Figure 9	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 22	I. Andersson, Department of Diagnostic Radiology, Malmö, Sweden	Figure 35	Hakama <i>et al.</i> (1999)
Figure 10	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 23	I. Andersson, Department of Diagnostic Radiology, Malmö, Sweden	Figure 36	Hakama <i>et al.</i> (1999)
Figure 11	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 24	I. Andersson, Department of Diagnostic Radiology, Malmö, Sweden	Figure 37	Working Group
Figure 12	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 25	Boyd <i>et al.</i> (2001)	Figure 38	Working Group
Figure 13	I. Ellis, Division of Histopathology, Nottingham City Hospital, UK	Figure 26	I. Andersson, Department of Diagnostic Radiology, Malmö, Sweden	Figure 39	Working Group
				Figure 40	European Network of Cancer Registries (2001)
				Figure 41	Working Group
				Figure 42	S.H. Heywang-Köbrunner, Clinic of Radiology, Martin Luther University, Halle, Germany
				Figure 43	National Cancer Institute (2001)
				Figure 44	de Koning <i>et al.</i> (1991)
				Figure 45	de Koning <i>et al.</i> (1991)
				Figure 46	Boer <i>et al.</i> (1995)
<hr/>					
Page 40	Clinical Breast Examination. JAMA (1999) 282, 1270–80 American Medical Association	Page 44	Breast Self examination. Sharon McLaren, University of Windsor, Canada	Page 66	Access Arts Inc., Brisbane Powerhouse, Australia
		Page 46	Lancet Oncology, 3, p. 247 April 2002		