International Agency for Research on Cancer

World Health Organization

IARC Handbooks of Cancer Prevention

PREAMBLE FOR SECONDARY PREVENTION

Lyon, France
October 2019

Contents

PREAMBLE – SECONDARY PREVENTION

The Preamble to the *IARC Handbooks of Cancer Prevention* describes the objectives and scope of the programme, general principles and procedures, and scientific review and evaluations. The *IARC Handbooks* embody the principles of scientific rigour, impartial evaluation, transparency, and consistency. The Preamble should be consulted when reading an *IARC Handbook* or a summary of an *IARC Handbook*'s evaluations. Separate Instructions for Authors describe the operational procedures for the preparation and publication of a volume of the *IARC Handbooks*.

A. GENERAL PRINCIPLES AND PROCEDURES

1. Background

Prevention of cancer is the mission of the International Agency for Research on Cancer (IARC). Cancer prevention is needed even more today than when IARC was established, in 1965, because the global burden of cancer is high and continues to increase, as a result of population growth and ageing and increases in cancer-causing exposures and behaviours, especially in low- and middle-income countries (Stewart & Kleihues, 2003; Boyle & Levin, 2008; Stewart & Wild, 2014).

Broadly defined, prevention is "actions aimed at eradicating, eliminating, or minimizing the impact of disease and disability, or if none of these is feasible, retarding the progress of disease and disability" (Porta, 2014). Cancer prevention encompasses primary, secondary, and tertiary prevention. Primary prevention consists of actions that can be taken to lower the risk of developing cancer. Secondary prevention entails methods that can find and ameliorate precancerous conditions or find cancers in the early stages, when they can be treated more successfully. Tertiary prevention is the application of measures aimed at reducing the impact of long-term disease and disability caused by cancer or its treatment.

The *IARC Handbooks of Cancer Prevention* provide critical reviews and evaluations of the scientific evidence on the preventive effects of primary or secondary cancer prevention measures. The evaluations of the *IARC Handbooks* are used by national and international health agencies to develop evidence-based interventions or recommendations for reducing cancer risk.

The *IARC Handbooks of Cancer Prevention* series was launched in 1995 by Dr Paul Kleihues, then Director of IARC, in recognition of the need for a series of publications that would critically review and evaluate the evidence on a wide range of cancer-preventive interventions. The first volume of the *IARC Handbooks* (IARC, 1997) reviewed the evidence on cancer-preventive effects of non-steroidal anti-inflammatory drugs, specifically aspirin, sulindac, piroxicam, and indomethacin. *Handbooks* Volume 6 (IARC, 2002a) was the first that evaluated behavioural interventions (weight control and physical activity), and *Handbooks* Volume 7 (IARC, 2002b) was the first that evaluated cancer screening (breast cancer screening). *Handbooks* Volumes 11–14 (IARC, 2007, 2008, 2009, 2011) focused on tobacco control. After a 3-year hiatus, the *IARC Handbooks* series was relaunched in 2014 with the preparation of *Handbooks* Volume 15 (IARC, 2016a), which re-evaluated breast cancer screening.

IARC's process for developing *Handbooks* engages international, expert scientific Working Groups in a transparent synthesis of different streams of evidence, which is then translated into an overall evaluation according to criteria that IARC has developed and refined (see Part A, Section 6). Scientific advances are periodically incorporated into the evaluation methodology, which must enable the evaluation of new generations of existing methods as well as new screening methodologies.

This Preamble, first prepared as the *Handbooks* Working Procedures in 1995 and later adapted to the topics of cancer screening and tobacco control, is primarily a statement of the general principles and procedures used in developing a *Handbook*, to promote transparency and consistency across *Handbooks* evaluations. In addition, IARC provides Instructions for Authors to specify more detailed operating procedures.

2. Objectives, scope, and definitions

2.1 Objectives and scope

The scope of the *IARC Handbooks of Cancer Prevention* series is to contribute to reducing the incidence of or mortality from cancer worldwide. To this end, the *IARC Handbooks* programme prepares and publishes, in the form of volumes of *Handbooks*, critical scientific reviews and evaluations of the available evidence on the efficacy, effectiveness, and harms of a wide range of cancer-preventive interventions. The primary target audiences for the *Handbooks* are national and international agencies with responsibility for, or advocating for, public health. The *IARC Handbooks* are an important part of the body of information on which public health decisions for cancer prevention may be based. However, public health options to prevent cancer vary from one setting to another and from country to country, and relate to many factors, including socioeconomic conditions and national priorities. Therefore, no recommendations are given in the *Handbooks* with regard to regulations or legislation, which are the responsibility of individual governments or other international authorities. However, the *IARC Handbooks* may aid national and international authorities in devising programmes of health promotion and cancer prevention, estimating the balance of benefits and harms, and considering cost–effectiveness evaluations.

The *IARC Handbooks* programme also does not make formal research recommendations. However, because *Handbooks* synthesize and integrate streams of evidence on cancer prevention, critical gaps in knowledge that merit research may be identified.

2.2 Definition of interventions for secondary prevention

The current *IARC Handbook* addresses a specific intervention or class of interventions for **secondary prevention**. The principal instruments of secondary prevention of cancer are interventions for early detection of precancerous lesions (i.e. precancer) or invasive cancer, which are currently mostly cancer screening interventions. However, there is growing evidence that action campaigns to increase awareness of cancer among the general public can increase the number of people who present to health-care providers, leading to earlier diagnosis of cancer and, generally, to better cancer outcomes. Such interventions for early diagnosis are also within the scope of the *Handbooks* programme.

Screening is the systematic application of a test that "can be applied rapidly in a presumably asymptomatic population, aiming at the presumptive identification of unrecognized disease or defect" (Porta, 2014). Screening tests sort out apparently-well people who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic, because people with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment (Porta, 2014). Screening may enable diagnosis of cancer sufficiently early that cure and resulting prevention of cancer death or a reduction in risk of cancer are realistic possibilities. Screening for some cancers, such as cervical cancer or colorectal cancer, may also detect precancer, effective treatment of which can prevent occurrence of invasive cancer. Screening can also cause harm, and evidence for harm must also be considered when evaluating the capacity of screening to reduce the incidence of cancer or death from cancer.

Screening interventions can be applied across a continuum of:

(i) the general population (often circumscribed by age and sex);

(ii) subgroups with particular predisposing host characteristics, such as genetic susceptibility, precursor lesions, or particular diseases other than cancer, or with high exposure to environmental, occupational, or behavioural risk factors; and

(iii) people with a history of cancer who are at high risk of a further primary cancer.

Early diagnosis interventions aim at detecting cancer in symptomatic patients as early as possible. Delays in accessing cancer care are common with late-stage presentation, particularly in lower-resource settings and in vulnerable populations. The consequences of delayed or inaccessible cancer care are lower likelihood of survival, greater morbidity of treatment, and higher costs of care, resulting in avoidable deaths and disability from cancer. Early diagnosis improves cancer outcomes by providing care at the earliest possible stage and

is therefore an important public health strategy in all settings (https://www.who.int/cancer/prevention/diagnosis-screening/en/). One of the most commonly used strategies is to raise awareness among the public and/or health professionals of early signs and symptoms of cancer in order to facilitate diagnosis before the disease becomes advanced. Other possible interventions to promote early diagnosis may involve regulation of health care and organization of health services (WHO, 2017).

2.3 Definitions of efficacy, effectiveness, and harms

Efficacy and effectiveness are two fundamental concepts underlying the evaluation of preventive interventions (Cochrane, 1972). Efficacy was defined by Porta (2008) as "the extent to which a specific intervention, procedure, regimen or service produces a beneficial result under ideal conditions … Ideally, the determination of efficacy is based on the results of a randomized controlled trial". Effectiveness was defined by Porta (2008) 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 and effectiveness of an intervention at the population level is an important one to make when evaluating preventive interventions. Efficacy is a necessary, but not sufficient, basis for formulating recommendations for an intervention. Whereas efficacy of an intervention can be inferred if effectiveness is established, efficacy does not guarantee effectiveness because of the number of implementation steps, each with uncertainty, required to deliver an efficacious prevention intervention as an effective programme in a target population. Ideally, efficacy is established before a preventive intervention is implemented in a whole community or population, so as to determine whether a case for population-wide implementation can be made on the basis of the balance of the benefits and harms and the financial costs of the intervention. However, it has not been unusual for preventive interventions to be implemented in the absence of evidence of efficacy. Should that occur, evaluation of effectiveness may be the only way to determine whether the case for the intervention is strong enough to justify its continuation or implementation elsewhere.

In addition to being shown to be efficacious or effective, screening interventions must satisfy other requirements if they are to be considered for implementation in practice, including an acceptable balance of benefits and harms. In the present context, harm is defined as any impairment or increase in risk of impairment as a result of exposure to or participation in a preventive intervention. Harms include physical, psychological, social, and economic consequences of a preventive intervention. Adverse events in health care are a subset of harms. Evaluation of these potential harms is an important component of the summary of the evidence.

For screening and for early diagnosis, other issues to be considered include acceptability to the target population, impact on health equity, cost, cost–effectiveness, availability of the personnel and facilities required to deliver the screening intervention, and access to the health services needed to diagnose and treat the disease detected. Depending on the specific intervention, some of these issues may be of sufficiently high interest to programme managers that they, too, are reviewed in the *IARC Handbook*.

Although the distinction between evidence of efficacy and effectiveness is an important one to make when seeking to act on cancer prevention, the *Handbooks* evaluations are based on evidence from **all** relevant research into efficacy and effectiveness.

3. Identification and selection of interventions and outcomes for review

3.1 Development of an analytical framework

As one of the first steps in the review and evaluation of a selected cancer screening intervention, the IARC Secretariat, with the support of the Working Group, drafts an analytical framework. Such a framework depicts the relationships among the study population, intervention, comparator, and intermediate outcomes or changes in health status as relevant. The analytical framework includes both benefits and harms, and key contextual issues related to participation and implementation of the intervention and its impact on population health. The framework defines the intervention in its broadest context and specifies the aspects for which the *Handbook* will review and evaluate the evidence.

In this framework, it is most commonly the case that a single cancer type, usually only topographically defined, is the primary target, and the reduction of the incidence of and/or mortality from that cancer type is the primary outcome. However, it is sometimes the case that intermediate outcomes (i.e. outcomes that are not invasive cancer or death from cancer) are important targets. For example, detection and ablation of precancerous polyps is the mechanism whereby some screening methods for colon cancer and rectal cancer reduce the incidence of colorectal cancer. Moreover, it is plausible that a new test with high sensitivity and specificity for a precancerous lesion, such as high-grade cervical intraepithelial neoplasia, could be judged on the grounds of these characteristics to be efficacious in preventing invasive cervical cancer and death from cervical cancer, provided that there is also strong evidence that ablation of the precancerous lesion prevents invasive cervical cancer. These possibilities are taken into consideration when defining the framework of a *Handbook.*

3.2 Selection of the interventions

For each new volume of the *Handbooks*, IARC selects one or more interventions for review by considering the availability of pertinent research studies, the need to evaluate an important development in cancer prevention, or the need to re-evaluate a previously evaluated intervention. IARC will also consider current public health priorities in specific geographical regions, for example the concerns of countries or regions with a high risk of specific cancer types (see Part A, Section 6, Step 1).

Interventions not previously evaluated in the *IARC Handbooks* series are selected for evaluation, where the body of evidence is large enough to warrant evaluation, on the basis of one or both of the following criteria:

• The intervention is of putative preventive value, but its effects or balance of benefits and harms have not been established formally;

• The available evidence suggests that the intervention has the potential to significantly reduce the incidence of or mortality from cancer, or to have a significant impact on an intermediate outcome (e.g. precancerous lesions; see below) known or highly suspected to be linked to cancer (see Part A, Section 6, Step 2).

In addition, an intervention previously evaluated in a *Handbook* may be re-evaluated if important new data become available about its effects, or if its technology or implementation has changed enough for there to be substantial changes in its effects. Occasionally, a re-evaluation may be limited to specific aspects of the screening intervention to which the new evidence predominantly relates (e.g. tomosynthesis for breast cancer screening). For re-evaluations, the full body of evidence relevant to the intervention of interest is considered, either by de novo review of all evidence or by accepting as accurate the evidence review of the previously published *Handbook* and undertaking a de novo review of evidence published since the previous review. Both approaches lead to an evaluation based on all relevant evidence (see Part A, Section 6, Steps 4 and 5). The choice of the approach is subject to the judgement of the Working Group.

4. The Working Group and other meeting participants

Five categories of participants can be present at *IARC Handbooks* meetings (Table 1):

(i) *Working Group* members have ultimate responsibility for determining the final list of studies that contribute evidence to the evaluation, performing the scientific review of the evidence, and making the final, formal evaluation of the strength of evidence for the capacity of the screening interventions to reduce cancer incidence or cancer mortality. The Working Group is multidisciplinary and is organized into Subgroups of experts in the fields that the *Handbook* covers.

IARC selects the Working Group members on the basis of relevant expertise and an assessment of declared interests (see Part A, Section 5). For screening, the fields of expertise are: (i) the cancer targeted and its global epidemiology; (ii) worldwide use of preventive interventions for the cancer targeted; and (iii) specific knowledge and experience of screening, in general or as practised for the targeted cancer. Consideration is also given to

diversity in scientific approaches, in stated positions on the strength of the evidence supporting the intervention, and in demographic characteristics. Working Group members generally have published research related to the interventions being reviewed or to the cancer types or intermediate outcomes that the interventions being reviewed are thought to prevent; IARC uses literature searches to identify most experts. IARC also encourages public nominations through its Call for Experts. IARC's reliance on Working Group members with expertise on the subject matter or relevant methodologies is supported by decades of experience documenting that there is value in specialized expertise and that the overwhelming majority of Working Group members are committed to the objective evaluation of scientific evidence and not to the narrow advancement of their own research results or a predetermined outcome (Wild & Cogliano, 2011). Working Group members are expected to serve the public health mission of IARC and to refrain from using inside information from the meeting or meeting drafts for financial gain until the full volume of the *Handbooks* is published (see also Part A, Section 7).

IARC selects, from among the Working Group members, individuals to serve as Meeting Chair and Subgroup Chairs. Subgroup Chairs have preferably served in previous *Handbooks* meetings as Working Group members or in similar review processes. At the opening of the meeting, the Working Group is asked to endorse the Meeting Chair selected by IARC or to propose an alternative. The Meeting Chair and Subgroup Chairs take a leading role at all stages of the review process (see Part A, Section 7) to promote open scientific discussions that involve all Working Group members in accordance with committee procedures and to ensure adherence to the processes described in this Preamble.

(ii) *Invited Specialists* are experts with critical knowledge and experience on the interventions being reviewed, the cancer types that the interventions being reviewed are thought to prevent, or relevant methodologies, but who have a declared conflict of interests that warrants exclusion from developing or influencing the evaluations. The Invited Specialists do not draft any section of the *Handbook* that pertains to the description or interpretation of the data on which the evaluation is based, or participate in the evaluations. Invited Specialists are invited in limited numbers, when necessary, to assist the Working Group by contributing their unique knowledge and experience to the discussions.

(iii) *Representatives of national and international health agencies* may attend because their agencies are interested in the subject of the *Handbook*. The Representatives of national and international health agencies do not draft any section of the *Handbook* or participate in the evaluations. Representatives can participate in discussions at times designated by the Meeting Chair or a Subgroup Chair. Relevant World Health Organization (WHO) staff members attend as members of the *IARC Secretariat* (see below).

(iv) *Observers* with relevant scientific credentials are admitted in limited numbers. Attention is given to the balance of Observers from entities with differing perspectives on the interventions under review. Observers are invited only to observe the meeting, do not draft any section of the *Handbook* or participate in the evaluations, must agree to respect the Guidelines for Observers at *IARC Handbooks* meetings (IARC, 2018), and must not attempt to influence the outcomes of the meeting. Observers may speak at Working Group or Subgroup sessions at the discretion of the Chair.

(v) The *IARC Secretariat* consists of scientists who are designated by IARC or WHO and who have relevant expertise. The IARC Secretariat coordinates and facilitates all aspects of the review and evaluation process and ensures adherence to the processes described in this Preamble throughout the development of the scientific reviews and evaluations (see Part A, Sections 5 and 6). The IARC Secretariat announces and organizes the meeting, identifies and invites the Working Group members, and assesses the declared interests of all meeting participants in accordance with WHO requirements (see Part A, Section 5). The IARC Secretariat supports the activities of the Working Group (see Part A, Section 7) by performing systematic literature searches, performing title and abstract screening, organizing conference

calls to coordinate the development of drafts and to discuss cross-cutting issues, and reviewing drafts before and during the meeting. Members of the IARC Secretariat serve as meeting rapporteurs, assist the Meeting Chair and Subgroup Chairs in facilitating all discussions, and may draft text or tables or assist a Subgroup in the conduct of additional analyses when designated by the Meeting Chair or a Subgroup Chair. After the meeting, the IARC Secretariat reviews the drafts for factual accuracy of research results cited. The participation of the IARC Secretariat in the evaluations is restricted to clarifying or interpreting the Preamble.

All meeting participants are listed, with their principal affiliations, in the front matter of the published volume of the *Handbooks*. Pertinent interests, if any, are listed in a footnote to the participant's name. Working Group members and Invited Specialists serve as individual scientists and not as representatives of any organization, government, or industry (Cogliano et al., 2004).

The roles of the participants are summarized in Table 1.

Table 1. Roles of participants at IARC *Handbooks* meetings

Category of participant	Role			
	Prepare text, tables, and analyses	**Participate in discussions**	**Participate in evaluations**	**Eligible to serve as Meeting Chair or Subgroup Chair**
Working Group members	√	√	√	√
Invited Specialists	√[a]	√		
Representatives of health agencies		√[b]		
Observers		√[b]		
IARC Secretariat	√[c]	√	√[d]	

[a] Only for sections not directly relevant to the evaluation

[b] Only at times designated by the Meeting Chair and/or Subgroup Chair

[c] Only when needed or requested by the Meeting Chair and/or Subgroup Chair

[d] Only for supporting Working Group members and for clarifying or interpreting the Preamble

5. Development of a volume of the *IARC Handbooks*

Each volume of the *Handbooks* is developed by an ad hoc, specifically convened Working Group of international experts. Approximately 1 year before the meeting of a Working Group, a preliminary list of interventions to be reviewed (see Part A, Section 3), together with a Call for Data and a Call for Experts, is announced on the *Handbooks* programme website (http://handbooks.iarc.fr/).

The IARC Secretariat selects potential Working Group members based on the criteria described in Part A, Section 4. Before a meeting invitation is extended, each potential participant, including the IARC Secretariat, completes the WHO Declaration of Interests form to report financial interests, employment and consulting (including remuneration for serving as an expert witness), individual and institutional research support, and non-financial interests such as public statements and positions related to the subject of the meeting. IARC assesses the declared interests to determine whether there is a conflict that warrants any limitation on participation (see Table 1).

Approximately 2 months before a meeting, IARC publishes on the *Handbooks* programme website the names and principal affiliations of all participants and discloses any pertinent and significant conflicts of interests, for transparency and to provide an opportunity for undeclared conflicts of interests to be brought to IARC's attention. It is not acceptable for Observers or third parties to contact other participants before a meeting or to lobby them at any time. Meeting participants are asked to report all such contacts to IARC (Cogliano et al., 2005).

The Working Group meets at IARC to discuss and finalize the scientific review and to develop summaries and evaluations. At the opening of the meeting, all meeting participants update their Declarations of Interests forms, which are then reviewed for conflicts of interest by IARC. Declared interests related to the subject of the meeting are disclosed to the meeting participants during the meeting and in the published volume of the *Handbooks* (Cogliano et al., 2004).

The objectives of the meeting are twofold: peer review of the drafts and consensus on the evaluations. During the first part of the meeting, Working Group members work in Subgroups to review the pre-meeting drafts, develop a joint Subgroup draft, and draft Subgroup summaries. During the last part of the meeting, the Working Group meets in plenary sessions to review the Subgroup drafts and summaries and to develop the consensus evaluations. As a result, the entire volume is the joint product of the Working Group and there are no individually authored sections. After the meeting, the master copy is verified by the IARC Secretariat (see Part A, Section 4(v)), edited, and prepared for publication. The aim is to publish the volume of the *Handbooks* within approximately 12 months of the Working Group meeting. The IARC Secretariat prepares a summary of the outcome for publication in a scientific journal or on the *Handbooks* programme website soon after the meeting.

The time frame and milestones for public engagement during the development of a volume of the *IARC Handbooks* are summarized in Table 2.

Table 2. Public engagement during the development of a volume of the *IARC Handbooks*

Approximate time frame	Milestones
~1 year before a *Handbooks* meeting	IARC posts on the *Handbooks* programme website:
	Preliminary List of Interventions to be reviewed
	Call for Data and Call for Experts open
	Requests for Observer Status open
	WHO Declarations of Interests form
~8 months before a *Handbooks* meeting	Call for Experts closes
~4 months before a *Handbooks* meeting	Requests for Observer Status close
~2 months before a *Handbooks* meeting	IARC publishes the names, principal affiliations, and declared conflicts of interest of all meeting participants, and a statement discouraging contact of Working Group members by outside parties
~1 month before a *Handbooks* meeting	Call for Data closes
Handbooks **meeting**	
~2–4 months after a *Handbooks* meeting	IARC publishes a summary of evaluations and key supporting evidence as a scientific article in a high-impact journal or on the *Handbooks* programme website
~9–12 months after a *Handbooks* meeting	IARC Secretariat publishes the verified and edited master copy of the plenary drafts as a *Handbooks* volume

6. Overview of the scientific review and evaluation process

Principles of systematic review are applied to the identification, screening, synthesis, and evaluation of the evidence (as described in Part B, Sections 2–7 and detailed in the Instructions for Authors). For each volume of the *Handbooks*, the information on the conduct of the literature searches, including search terms and the inclusion and exclusion criteria that were used for each relevant stream of evidence, is recorded.

The Working Group considers all relevant studies, including experimental and observational studies of the efficacy and/or effectiveness of the intervention and related harms (including systematic reviews and meta-analyses), pertinent information on global practices of the screening methods, and background information on the global epidemiology and burden of the targeted cancer type.

In general, only studies that have been published or accepted for publication in the openly available scientific literature are reviewed. Materials that are publicly available and whose content is final may be reviewed if there is sufficient information to enable peer evaluation of the quality of the methods and results of the studies (see Step 1, below). Such material may include reports from government agencies, dissertations for higher degrees, and other apparently reputable scientific sources. Systematic Internet searches for potentially relevant "grey literature" are not usually done. The reliance on published and publicly available studies promotes transparency and protects against citation of information that, although purportedly final, may change before it is published.

The steps of the review process are as follows:

Step 1. Identification of the review question: After the intervention (or interventions) and outcome (or outcomes) to be reviewed have been specified, the IARC Secretariat, in consultation with the Working Group, drafts the review question (or questions) in PICO form (population, intervention/exposure, comparator, and outcome) as required to determine the inclusion and exclusion criteria for the studies. An analytical framework is developed to assist in identifying and formulating the review questions, with the aim of making as large a contribution as possible to the global prevention of cancer.

Step 2. Comprehensive and transparent identification of the relevant information: The IARC Secretariat specifies search terms for the key PICO components of each question and identifies relevant studies through initial comprehensive literature searches in authoritative biomedical databases (e.g. PubMed). The literature searches are designed in consultation with a librarian and other technical experts. The scope and specifications of the searches may be modified, and the searches rerun, depending on the amount, relevance, and perceived completeness of the articles they identify. The IARC Secretariat may also identify relevant studies from reference lists of past *Handbooks*, retrieved articles, or authoritative reviews, and through the Call for Data (see Table 2). The Working Group provides input and advice to the IARC Secretariat to refine the search strategies, and identifies additional articles through other searches and personal expert knowledge.

For certain types of interventions (e.g. administration of regulated imaging agents), IARC also gives relevant regulatory authorities, and parties regulated by such authorities, an opportunity to make pertinent unpublished studies publicly available by the date specified in the Call for Data. Consideration of such studies by the Working Group is dependent on the public availability of sufficient information to enable an independent peer evaluation of: (i) completeness of reporting of pertinent data; (ii) study quality; and (iii) study results.

Step 3. Screening, selection, and organization of the studies: The IARC Secretariat screens the retrieved articles by reviewing the title and abstract against the inclusion and exclusion criteria agreed upon by the Working Group and technical experts in the review process. Potentially relevant studies are then made available to Working Group members for full-text screening and inclusion in or exclusion from the evidence base using agreed criteria specific to this task.

Step 4. Extraction of information from included studies, including characteristics relevant to study quality: Working Group members, working individually as members of defined Subgroups before the *Handbooks* meeting, review and succinctly describe pertinent characteristics and results of included studies as detailed in Part B, Sections 2–5. Study design and results are tabulated systematically in a standard format. This step may be iterative with Step 5.

11

Step 5. Assessment of study quality: Also before the *Handbooks* meeting, Working Group members evaluate the quality and informativeness of each study they included based on the considerations (e.g. design, conduct, analysis, and reporting of results) described in Part B, Sections 2–5. Evaluation of study quality can be done either narratively or by use of a risk of bias assessment tool when a relevant one is available and can add value to the process. Interpretations of the results, and the strengths and limitations of each study, are clearly outlined in square brackets as part of the description of that study (see Part B).

Step 6. Peer review: Several months before the meeting, the pre-meeting drafts produced from Steps 4 and 5 are peer-reviewed by other members of the Working Group (usually within the same Subgroup). The IARC Secretariat also reviews the drafts for completeness, consistency between drafts, and adherence to the *Handbooks* Instructions for Authors. The peer-review comments are sent to the Working Group members, who produce a revised pre-meeting draft. The revised drafts are reviewed and revised in Subgroup sessions during the *Handbooks* meeting.

Step 7. Synthesis of results and quality of the studies: The results and quality of the included studies are synthesized by the Working Group to provide a summary of the evidence and its quality for each outcome. This synthesis can be narrative or quantitative (for details, see the Instructions for Authors), and the quality synthesis may include use of an overall quality of evidence assessment tool, such as GRADE (Siemieniuk & Guyatt, 2019).

Meta-analyses of large bodies of evidence may be performed by the Working Group and/or by the IARC Secretariat before the meeting if such meta-analyses would assist in evidence synthesis and evaluation. For more information on the conduct and use of such meta-analyses, see Part B, Section 5.1c.

Step 8. Interpretation of study results and evaluation of strength of evidence: The whole Working Group reviews the study descriptions and the summaries of the body of evidence for each outcome or end-point, discusses the overall strengths and limitations of the evidence in each stream of data, and evaluates the strength of evidence for a preventive effect on cancer or an intermediate outcome in each stream using transparent methods, which may include the use of established specific tools. The preventive effect for each stream of evidence is assessed. The Working Group then integrates the assessments from all streams of evidence (see Part B, Section 7.1) and develops the rationale for its consensus evaluation of the preventive effect of the screening or early diagnosis method (see Part B, Section 7.2).

7. Responsibilities of the Working Group

The Working Group is responsible for the final list of studies included in the evaluation and the review and evaluation of the evidence for a *Handbook*, as described above. The IARC Secretariat supports these activities (see Part A, Section 4). To ensure that the process is rigorous, independent, and free from individual conflicts of interest, Working Group members must accept the following responsibilities:

(i) Before the meeting, Working Group members:

• help in developing the analytical framework;

• ascertain that all appropriate studies have been identified and selected;

• assess the methods and quality of each included study;

• prepare pre-meeting drafts that present an accurate quantitative and/or textual synthesis of the body of evidence, with key elements of study design and results and notable strengths and limitations;

• participate in conference calls organized by the IARC Secretariat to coordinate the development of pre-meeting drafts and to discuss cross-cutting issues; and

• review and provide comments on pre-meeting drafts prepared by other members of their Subgroup or of the Working Group.

(ii) At the meeting, Working Group members work in Subgroups to:

• critically review, discuss, and revise the pre-meeting drafts and adopt the revised versions as consensus Subgroup drafts; and

• develop and propose an evaluation of the strength of the evidence summarized in the consensus Subgroup drafts (see Part B, Section 6), using the *IARC Handbooks* criteria (see Part B, Section 7.1).

(iii) At the meeting, Working Group members work in plenary sessions to:

• present their Subgroup drafts for scientific review by and discussion with the other Working Group members, and subsequent revisions, as needed;

• participate in review and discussion of other Subgroup drafts and in their adoption as a consensus Working Group draft;

• participate in review and discussion of the summaries and evaluations of the strength of the evidence developed in Subgroups (see Part B, Section 7.1), and contribute to their revision, as needed, and their adoption by consensus of the full Working Group; and

• contribute to the discussion of and adoption by consensus of an overall evaluation proposed by the Meeting Chair using the guidance provided in Part B, Section 7.1.

The Working Group strives to achieve consensus evaluations. Consensus reflects broad agreement among the Working Group members, but not necessarily unanimity. If unanimity has not been reached when the interpretations of the evidence by all Working Group members have been expressed and debated, the judgement of the majority of the Working Group members is taken as the consensus. When consensus is reached in this way, the Meeting Chair may poll Working Group members to determine and record the diversity of scientific opinion on the overall evaluation.

Only the final product of the plenary sessions represents the views and expert opinions of the Working Group. The *Handbook* is the joint product of the Working Group and represents an extensive and thorough peer review of the body of evidence (review of individual studies, synthesis, and evaluation) by a multidisciplinary group of experts. Initial pre-meeting drafts and subsequent revisions are temporarily archived but are not released, because they would give an incomplete and possibly misleading impression of the consensus developed by the Working Group over its complete deliberation.

B. SCIENTIFIC REVIEW AND EVALUATION

This part of the Preamble discusses the types of evidence that are considered and summarized in each section of a *Handbook*, followed by the scientific criteria that guide the evaluations. In addition, a section of General Remarks at the front of the volume discusses the reasons the interventions were scheduled for evaluation and any key issues encountered during the meeting.

1. Definitions

Secondary prevention of cancer is the use of methods that can lead to the detection of asymptomatic or early symptomatic precancerous conditions or cancers at a stage when treatment of a lesion that is found can prevent progression to invasive cancer or, if the cancer is already invasive, prevent death from cancer. The two cornerstones of secondary prevention are screening and early diagnosis. WHO defines these terms as follows (https://www.who.int/cancer/prevention/diagnosis-screening/en/).

Screening is "the systematic application of a screening test in a presumably asymptomatic population. It aims to identify individuals with an abnormality suggestive of a specific cancer. These individuals require further investigation."

Early diagnosis is "the early identification of cancer in patients who have symptoms of the disease". Early diagnosis is most commonly achieved by raising "the awareness (by the public or health professionals) of early signs and symptoms of cancer in order to facilitate diagnosis before the disease becomes advanced. This enables more effective and simpler therapy."

WHO defines a cancer *early detection programme* as "the organized and systematic implementation of early diagnosis or screening (or both), diagnosis, treatment, and follow-up", thus encompassing both screening and early diagnosis. Early detection programmes, when implemented, usually operate alongside opportunistic early diagnosis and/or screening.

IARC defines an ***organized screening programme*** as one that has "an explicit policy with specified age categories, method, and interval for screening; a defined target population; a management team responsible for implementation; a health-care team for decisions and care; a quality assurance structure; and a method for identifying cancer occurrence in the target population" (IARC, 2005). In principle, an organized screening programme also includes systematic invitation of the target population for quality-assured screening tests and assured follow-up of screen-positive subjects with diagnostic investigations, treatment, and post-treatment care. The former can minimize inequalities in access to screening by giving every eligible and contactable person access to screening.

Opportunistic refers to the fact that the medical examination is requested by a patient or offered by a health practitioner in the context of the patient–practitioner relationship and is not, or is minimally, subject to any other organizing principle. The proportion of screening for a particular cancer that is opportunistic varies widely from country to country; in many countries screening is exclusively opportunistic, and in some countries screening is almost exclusively organized (for particular types of cancer).

Compared with opportunistic screening, organized screening focuses much greater attention on higher coverage by way of systematic invitation and on the quality of the screening process, and provides greater protection against the harms of screening, including overscreening, poor-quality screening, adverse events of screening, and poor follow-up of those who test positive (Miles et al., 2004). The *IARC Handbooks* assess all available relevant evidence from both organized programmes and opportunistic settings in their evaluation of the effectiveness of a screening method or early diagnosis method.

Whether organized or opportunistic, screening is a complex public health strategy that requires substantial health-care resources, infrastructure, and coordination to be effective. In addition, screening should be undertaken only when efficacy and, ideally, effectiveness have been established. It should also only be undertaken when resources are sufficient to cover a large proportion of the intended target group, when facilities exist for follow-up of screen-positive subjects to confirm or exclude disease and ensure treatment, and where the disease is a sufficiently burdensome public health problem to justify the effort and costs of screening. In addition, information systems are essential to monitor inputs and evaluate outcomes.

Early diagnosis programmes of cancer also have minimum requirements, specifically the facilities needed to confirm or exclude a diagnosis of cancer in people who present to health-care providers with symptoms suggestive of a potentially curable cancer, and to ensure treatment when a diagnosis of cancer is confirmed. At present, the tools of early diagnosis are largely limited to community education about symptoms that may suggest cancer, and to educating or enabling primary care practitioners to ask at-risk patients presenting for *any* care about symptoms they have that may be signs of cancer. Evidence of the effectiveness of such measures is accumulating (Emery et al., 2014). Other possible interventions to promote early diagnosis may involve regulation of health care and organization of health services.

It is important to note that in low- and middle-income countries, depending on societal prioritization, early diagnosis programmes may be the only affordable option for increasing the detection of cancer when it is potentially curable. Screening (organized or opportunistic) may be unaffordable, although simulation of realistic cost–effectiveness (taking into account all societal costs) might make some programmes attractive.

Early diagnosis and screening are the early parts of a multistep process. The *Handbooks* consider for evaluation the methods used for early diagnosis and screening, and not the steps that follow in the process. Although the following details about the scientific review and evaluation refer specifically to screening interventions, they will also apply for the evaluation of early diagnosis interventions, with some adaptation as needed.

2. Characterization of the disease

This type of *Handbook* addresses screening for cancer at one specific site. Information is presented on the precursor or invasive lesions that cancer screening aims to detect. Each cancer or other lesion is precisely defined as to its location and morphology, using the appropriate codes from the latest International Classification of Diseases for Oncology (IARC, 2019a) and brief pathological criteria for its

diagnosis as published by IARC (IARC, 2019b). The global distribution and burden of the cancer are summarized, including regional differences, time trends, and credible projections of incidence and/or mortality, based on IARC's data from cancer registries. The natural history of the cancer and its established risk factors and preventive factors are briefly described. The nature and efficacy of evidence-based, potentially curative therapy is also briefly described, together with geographical variation in its nature and accessibility worldwide.

3. Screening methods

Screening methods for the relevant cancer site are considered for evaluation if they have been subject to one or more well-conducted randomized controlled trials with cancer incidence and/or mortality (see Part A, Section 3) as the trial outcome. Screening methods for which no randomized controlled trials are available may be evaluated if the body of evidence from observational studies is sufficiently large to warrant evaluation, especially for screening methods that are already in use in the community.

New screening methods and innovations in existing methods that may offer significant improvements in screening performance, increases in acceptability of screening, or reductions in cost of screening but that did not meet the threshold for detailed review and evaluation described above (i.e. are materially different from other methods under consideration and have not been subject to one or more well-conducted randomized controlled trials or are in widespread use in some countries), or for which the body of evidence was too limited to enable an evaluation to be performed, are also reviewed. The review includes a description and critical assessment of any studies on the performance or the screening effect of these new methods or innovations of existing methods.

Emerging methods may be evaluated in the absence of studies of efficacy or effectiveness if comparative data with an established screening method are available. Such comparative data may include data

(i) on performance against validated reference standards (including those of the International Organization for Standardization [ISO] when relevant);

(ii) on other performance characteristics in populations at average risk; and

(iii) on intermediate outcomes that provide data on efficacy or effectiveness (e.g. sensitivity, specificity, and interval cancer rate) (Young et al., 2016).

Ideally, such comparisons will have been made under conditions in which potential biases have been minimized. Possible differences in other important characteristics, such as acceptability and possibility of harm, are also taken into account.

Each method considered for evaluation is described, and its **state-of-the-art application** is outlined. The description of each method should include whether the goal of screening is to reduce cancer-specific mortality by primarily detecting invasive lesions, or to reduce cancer-specific incidence by primarily detecting precursor lesions. The characteristics of the target population, such as age ranges and sex, should be stated. Other relevant issues for the method should be addressed, including:

- equipment and training required;
- technical quality control;
- the screening protocol and its expected performance, including sensitivity and specificity;
- host factors that affect screening performance;
- any assessment protocol for screen-positive subjects; and
- quality assurance.

4. Current global screening practices

A brief overview of relevant screening practices in different regions of the world is presented, limiting the description to those countries or settings where screening takes place. The following aspects are summarized if available:

• policies and guidelines for, and regulation of, screening;

• the type of screening offered (e.g. opportunistic screening, organized population-wide programme);

• the screening methods most commonly used or recommended; and

• availability of facilities, extent of population coverage, and participation rates.

In addition, demographic, cultural, and behavioural considerations that affect participation in screening are presented in a global perspective, with some specific, local characteristics, as appropriate.

5. Epidemiological studies of each screening method

The evaluative processes described here are repeated in full, as far as they apply, for each screening method reviewed.

Relevant studies of cancer in humans are identified using systematic review principles, as described in Part A and further detailed in the Instructions for Authors provided to each Working Group. Eligible studies include: all studies in humans of the association of the screening intervention of interest with its cancer incidence, mortality, or intermediate outcome target (studies of benign neoplasms, pre-neoplastic lesions, and other outcomes are reviewed when they are outcomes sought by, or intermediate outcomes related to, the screening intervention reviewed); studies dealing with the accuracy (sensitivity, specificity, and predictive values) of the screening intervention; studies examining a putative harm as an outcome of the screening intervention; reports on the balance of benefits and harms of screening; and reports on the cost–effectiveness of screening. Search strategies must take into account the possibility that any of the above-mentioned outputs from a single study may have been published separately from the other outputs of the study. Multiple publications may arise from successive follow-ups of a single trial population or cohort, from analyses focused on different aspects of a screening–outcome association, or from inclusion of overlapping populations. In these situations, only the most recent publication or the one that provides the most, or most relevant, information should be included, unless circumstances warrant otherwise.

5.1 Evaluation of the preventive and harmful effects of the intervention

(a) Types of studies considered

Several types of epidemiological studies contribute to the evaluation of the benefits and harms of cancer screening. Benefits are the principal focus of this section.

(i) *Experimental studies:* Allocation by the investigator of the participants to the intervention (screening) or control condition, ideally by a random and blind process (to the investigator and the participant), is the defining characteristic of experimental studies. These studies can include classic individually randomized controlled trials, cluster-randomized controlled trials that include sufficient clusters to minimize probability of bias, and a range of other designs in which there is non-random allocation of participants to the intervention or control condition or there are too few randomization units to minimize bias.

In principle, experimental studies can provide evidence for efficacy or effectiveness of an intervention that is at low risk of bias. In particular, pragmatic trials (trials designed to test the effectiveness of the intervention in a broad routine clinical practice) can provide evidence of effectiveness when conducted in settings with populations at average risk.

Studies with a tandem design (i.e. the same population is screened with both methods consecutively) can also be useful, to assess an emerging method and its relative impact on screening outcomes.

(ii) *Observational studies:* Typically, observational studies include cohort studies (including variants such as case–cohort and nested case–control studies), case–control studies, cross-sectional studies, and ecological studies, all with cancer incidence or mortality as an outcome. In addition to these designs, innovations in epidemiology enable many variant designs that may be considered in *Handbooks* evaluations. Observational studies generally provide evidence of effectiveness only.

Cohort and case–control studies of screening typically relate individual exposure to the screening intervention under study to the incidence of or mortality from the target cancer in individuals, and provide an estimate of the relative incidence of or mortality from cancer as the main measure of screening effect. In addition, cross-sectional studies may be used to measure accuracy, such as sensitivity, specificity, and predictive values.

In ecological studies, the unit of investigation is not an individual but a whole population or a set of subgroups of a population, and cancer incidence or mortality is related to a summary measure of the exposure (screening method) of the whole population at different times, or aggregate measures of the exposure in the subgroups at the same time. Time-based ecological studies may be of particular interest in evaluating the impact of screening methods, because changes in cancer incidence or mortality, or harms, over interrupted time periods can be related to exposure to the screening method within a single population. Nevertheless, results from ecological studies should be interpreted with caution for two reasons: (i) because they are prone to misclassification of exposure within individual time or population units, due to the lack of individual data on exposure or outcome, and (ii) because of the limited ability to adjust for confounders. Therefore, ecological studies should generally be used to raise hypotheses and to support the evidence of results from experimental or other observational studies.

(b) Study quality and informativeness

The following paragraphs outline the general principles of description, analysis, and interpretation of epidemiological studies in a cancer screening context. It is important to note that the evaluation of cancer screening studies involves complexities that are uncommon to other fields of epidemiology. Some examples of these complexities are self-selection for screening, heterogeneity of opportunity to be screened, confounding with differential treatment, and the complexities of lead time, length sampling, and overdiagnosis (IARC, 2016b).

Epidemiological studies are susceptible to several different sources of error. Study quality is assessed as part of the structured expert review process undertaken by the Working Group. A key aspect of quality assessment is consideration of the possible roles of chance and bias in the interpretation of epidemiological studies.

Chance, also called "random variation", can produce misleading study results. This variability in study results is strongly influenced by the sample size: smaller studies are more likely than larger studies to have effect estimates that are imprecise and, therefore, are more likely to be misleading. Confidence intervals around a study's point estimate of effect are routinely used to indicate the range of values of the estimate that could be produced by chance. Both experimental and observational epidemiological studies are prone to effects of chance.

Bias is the effect of factors in study design, conduct, or reporting that lead an association to erroneously appear stronger than, weaker than, or opposite in direction to the association that really exists between an exposure and an outcome. Biases that require consideration are varied and can be broadly categorized as selection bias, information bias (e.g. screening intervention and outcome measurement error), and confounding bias (Rothman et al., 2008). Selection bias in an epidemiological study can occur when the inclusion of participants from the eligible population or their follow-up in the study is influenced by their exposure (screening use) or their outcome (usually disease occurrence). Under these conditions, the measure of association found or not found in the study may not accurately reflect the association or lack thereof that might otherwise have been found in the eligible population (Hernán et al., 2004). Information bias results from inaccuracy in intervention or outcome measurement. Both can cause an association between hypothesized cause and effect to appear stronger or weaker than it really is. Confounding arises when a third factor is associated with both the intervention and the outcome and,

because of this, influences the apparent association between them (Rothman et al., 2008). An association between the purportedly preventive intervention and another factor that is associated with an increase or a decrease in the incidence of or mortality from the disease can lead to a spurious association or the absence of a real association of the purportedly preventive intervention with the disease. When either of these occurs, confounding is present.

In principle, experimental studies are less prone to each of these sources of bias, because selection for intervention or non-intervention is determined by the investigator (usually by random allocation) and not by the study participants or their characteristics. However, bias may arise because of lack of concealment, non-random allocation, lack of blinding, post-randomization exclusions, or non-acceptance of or non-adherence by the study participants to the conditions of the study arm (screening or not screening) to which they were randomized when, as is usual in experimental studies of cancer screening, they are not blind to their study arm. In addition, even when they are blind to the study arm, a high degree of participant non-adherence may cause important information bias and potential confounding with variables related to the choice of whether to adhere or not adhere to the study conditions. Because of such possibilities for confounding, it is common practice to include key confounding variables in the data collected from or about participants, to enable statistical control of confounding.

Two other sources of bias may have important effects on the estimates of the screening efficacy: lead-time bias and length bias (Cole and Morrison, 1980; IARC, 2016b). Lead time is the period between screen detection and when a tumour would have been clinically diagnosed in the absence of screening. The survival time, defined as the time from the date of diagnosis of cancer to the date of death, of screen-detected cases is overestimated because of this lead time, even for individuals who do not benefit from screening. Therefore, lead-time bias can produce data that appear to support a favourable effect of screening, if conclusions are based on survival analysis.

The other important bias is length bias (or length-sampling bias). The probability of a tumour being detected at screening depends, at least in part, on its growth rate, because slow-growing tumours have a longer preclinical detectable phase compared with fast-growing tumours. Thus, tumours detected at screening are a biased sample of preclinical lesions, weighted towards slower-growing tumours, which are generally thought to be associated with a better prognosis and therefore longer survival. This again leads to bias apparently in favour of screening.

In assessing the quality of the studies, the Working Group considers the following aspects:

• **Study description:** Clarity in describing the study design and its implementation, and the completeness of reporting of all other key information about the study and its results.

• **Study population:** Whether the study population was appropriate for evaluating the association between the screening intervention and cancer. Whether the study was designed and conducted in a manner that would minimize selection bias and other forms of bias. The designated outcomes in the study population must have been identified in a way that was independent of the screening intervention, for both experimental studies and observational studies, and the screening intervention must have been assessed in a way that was not related to disease (outcome) status. In these respects, completeness of recruitment into the study from the population of interest and completeness of follow-up for the outcome (see below) are very important.

• **Outcome measurement:** The appropriateness of the outcome measure (incidence of cancer, mortality from cancer, or an intermediate outcome, as defined in Part B, Section 1) for the screening intervention and the cancer type under consideration, the outcome ascertainment methodology, and the extent to which outcome misclassification may have led to bias in the measure or measures of association (e.g. because of systematic differences between exposed and unexposed people in the way in which the outcome was ascertained, and lack of blinding of ascertainment of cancer outcomes, which requires the exercise of human judgement).

• **Intervention measurement:** This includes (i) the adequacy (including the validity and the reliability) of the methods used to assess the intervention in observational studies, and adherence to the intervention condition in experimental studies, and (ii) the likelihood (and direction) of bias in the

measure or measures of association because of intervention measurement error or misclassification in observational studies and non-adherence to the intervention condition and cross-contamination of the non-intervention group in experimental studies (as described in Part B, Section 5.1).

• **Assessment of potential confounding:** The extent to which the authors took into account in the study design and analysis potentially confounding variables, including co-exposures, that could influence the occurrence of the outcome and may be related to the intervention of interest. Particular to screening interventions is the possibility that for a given stage, people with screen-detected cancers receive better treatment than those with symptom-detected cancers. Important sources of potential confounding by such variables should, where possible, have been addressed in the study design, such as by randomization, matching, or restriction, or in the analysis by statistical adjustment. In some instances, where direct information on confounders is unavailable, use of indirect methods to evaluate the potential impact of confounding on intervention–outcome associations is appropriate (e.g. Axelson & Steenland, 1988; Richardson et al., 2014).

• **Other potential sources of bias:** Each epidemiological study is unique in its study population, its design, its data collection, and, consequently, its potential biases. All possible sources of bias are considered for their possible impact on the results. Several sources of bias have important effects on the estimation of screening efficacy. The possibility of reporting bias (selective reporting of some results) should also be explored.

• **Statistical methodology:** The studies are evaluated for the adequacy of the statistical analysis methods used and their ability to obtain unbiased estimates of intervention–outcome associations, confidence intervals, and test statistics for the significance of measures of association. Appropriateness of methods used to address confounding, including adjusting for matching when necessary and avoiding treatment of probable mediating variables as confounders, is considered. Detailed analyses of cancer risks in relation to summary measures of intervention, such as cumulative exposure to the intervention, or temporal variables, such as age at first intervention or time since first intervention, are reviewed and summarized when available.

For the sake of economy and simplicity, this Preamble refers to the **list of possible sources of error** with the phrase "**chance, bias, and confounding**", but it should be recognized that this phrase encompasses a comprehensive set of concerns pertaining to study quality. These elements of study quality do not constitute and **should not be used as a formal checklist** of indicators of study quality. Rather, the assessment by the Working Group is reported in a narrative way, in the form of comments in square brackets. The **judgement of the experts is critical** in determining **how much weight to assign to different issues** when considering how all these potential sources of error should be integrated and how to rate the potential for error related to each. However, it is important that the **process** undertaken, including the weight given to various studies, be **replicable** and be described in a way that is **transparent** to readers.

• **Study informativeness:** The informativeness of a study is its ability to show a true preventive effect, if one exists, of the intervention on the outcome, and not to show an effect if one does not exist. Key determinants of informativeness include having a study population of sufficient size to obtain precise estimates of effect, sufficient elapsed time from intervention to measurement of outcome for an effect, if present, to be observable, presence of adequate intervention contrast, and relevant and well-defined time windows for intervention and outcome.

(c) Meta-analyses and pooled analyses

Independent epidemiological studies of the same intervention with a comparatively weak effect or small sample size may produce inconclusive results that are difficult to summarize. Combined analyses of data from multiple studies may increase the precision of estimates. There are two types of combined analysis: (i) meta-analysis, which involves combining summary statistics, such as relative risks from individual studies, and (ii) pooled analysis, which involves a pooled analysis of the raw data from the individual studies (Greenland & O'Rourke, 2008). There are also "umbrella reviews", systematic reviews of multiple meta-analyses, which may be evaluated by the Working Group.

The strengths of combined analyses are increased precision due to increased sample size and, in the case of pooled studies, the opportunity to better control for potential confounders and to explore interactions and modifying effects that may help to explain heterogeneity between studies. A disadvantage of combined analyses is the possible lack of comparability of results from various studies, because of differences in specification of the intervention or the outcome, population characteristics, subject recruitment, data collection procedures, methods of measurement, and effects of unmeasured covariates, which may differ among studies. These differences in study methods and quality can influence the results of both pooled analyses and meta-analyses.

Meta-analyses considered by the Working Group may include high-quality published meta-analyses, updates of such meta-analyses, and new meta-analyses. When published meta-analyses are considered by the Working Group, the conduct and reporting quality of the meta-analyses will be carefully assessed against prior expectations set with reference to items in checklists for published systematic reviews and meta-analyses, such as AMSTAR (AMSTAR, 2017) and/or PRISMA (Moher et al., 2009), with additional checks made of the alignment of the systematic review specifications with those required for the *Handbooks* evaluation, the completeness of coverage of articles relevant to the evaluation compared with those ultimately included in the meta-analysis, and the accuracy of extraction of required data from the results of the individual studies.

Subject to the judgement of the IARC Secretariat and in consultation with the Working Group, the updating of meta-analyses or the conduct of ad hoc meta-analyses may be performed by the Working Group and/or by the IARC Secretariat during preparation for a *Handbooks* meeting, when there are sufficient studies of an intervention–outcome association to aid the Working Group's assessment of the association. When results from both experimental and observational studies are available, any combined analyses should be conducted separately for experimental efficacy studies, experimental effectiveness studies, and observational studies, with consideration given to separate combined analyses of cohort and case–control studies, because of their different propensities to bias. The results of such ad hoc meta-analyses, which are specified in the text of the *Handbook* by presentation in square brackets, may come from the addition of the results of more recent studies to those of published meta-analyses or from de novo meta-analyses. Additional details on the conduct of such ad hoc meta-analyses are provided in the Instructions for Authors.

Irrespective of the source of the information for the meta-analyses and pooled analyses, the criteria for information quality applied are the same as those applied to individual studies. The sources of heterogeneity among the studies contributing to them are carefully considered and the possibility of publication bias evaluated.

(d) Evaluation of new technologies

It is important that a new screening test or method is evaluated before it replaces existing technology. New technology need not be subject to a full controlled trial of efficacy if it is similar enough to the old technology and if the old technology has been shown to reduce cancer incidence or cancer mortality. A new technology is considered similar enough if the method of screening is based on the same principles as the old technology and targets lesions with the same biology. In such instances, instead of a full controlled trial of efficacy, the following are required: (i) adequate analytical and clinical validity of the test in human subjects; (ii) cross-sectional evaluation of diagnostic accuracy of the new method for intermediate outcomes validated in randomized controlled trials or in tandem studies in a screening population at average risk (Young et al., 2016); and (iii) a prospective evaluation over more than one screening round of the comparative performance of the two methods, including participation, detection rates, false-positive rates, interval cancer rates, and the burden and harms of screening (Irwig et al., 2006; Young et al., 2016). In the absence of a reduction in risk of interval cancer, any increase in test sensitivity is probably due to an increase in overdiagnosis (see Section 5.2), which could make the new technology more harmful, rather than more beneficial, than the old technology. If the Working Group decides to make a full evaluation of a new screening method in comparison with an existing screening method that has been established to reduce the incidence of cancer or death from cancer, it does a full systematic review of research evidence relevant to this question, as described in Part A, Section 6.

(e) Considerations in assessing the body of epidemiological evidence

The ability of the body of epidemiological evidence to inform the Working Group about the efficacy or effectiveness of a screening intervention is related to both the quantity and the quality of the evidence. There is no formulaic answer to the question of how many studies are needed from which to draw inferences about the efficacy or effectiveness of a screening intervention, although more than a single study in a single population will almost always be needed.

Experimental and observational studies are to be considered. Randomized controlled trials typically provide the strongest evidence, but observational studies also provide valuable and timely information. For example, observational studies can be done for initial evaluation of proposed screening methods and for evaluation of their effectiveness after dissemination has occurred.

After the quality of individual epidemiological studies has been assessed and the informativeness of the various studies on the association between screening and cancer or an intermediate outcome has been evaluated, the body of evidence is assessed and a consensus scientific judgement is made about the strength of the evidence that the screening method under review reduces the incidence of cancer or death from cancer. In making its judgement, the Working Group considers several aspects of the body of evidence (e.g. Hill, 1965; Rothman et al., 2008; Vandenbroucke et al., 2016).

A strong association (e.g. a large relative risk or a relative risk that is well below 1.0) is more likely to be causal than a weak association, because it is harder for confounding or other biases to create a greater association than the one that is observed. However, it is recognized that estimates of effect of small magnitude do not imply lack of causality and may have a substantial impact on public health if the disease is common or if the screening intervention is highly feasible and/or widely applicable. Estimates of effects of small magnitude can also contribute useful information to the assessment of screening efficacy or effectiveness if the magnitude of the effect correlates with the level of screening intervention in populations that are differently exposed.

Associations that are consistently observed in several studies of the same design, in studies that use different epidemiological approaches, or under different circumstances of intervention are more likely to indicate screening efficacy or effectiveness than are isolated observations from single studies. If there are inconsistent results among investigations, possible reasons for such inconsistencies are sought (e.g. populations studied, intervention characteristics, measurements of outcomes, differences in study informativeness because of time since initiation of the intervention, screening participation), and their implications for the overall findings are assessed.

Results of studies that are judged to be of high quality and highly informative are given more weight than those of studies that are judged to be methodologically less sound or less informative.

Temporality of the association is also an essential consideration, that is, the intervention must precede the outcome, and by a time period that is sufficiently long for observation of a screening effect to be plausible.

5.2 Harms of screening

Potential harms to individuals that are linked to the screening method under review are also reviewed. Evidence of harm may come from any type of epidemiological study (see Section 5.1a) and may also be reported in studies separately from evidence on the benefits of screening using the same criteria as for preventive effects. Although the *IARC Handbooks* do not formally evaluate the harms associated with screening in the way that is done for the benefits, the review of the evidence of harms aims to be as complete, rigorous, and informative as it is for the evidence of beneficial effects.

Occurrence of screening harms is reviewed and described, and their potential impacts are discussed. The evaluation of harms includes: (i) estimates of rates of false-positive and false-negative findings, overdiagnosis, and overtreatment, which are harms shared by all screening methods; and (ii) estimates of risks of harm intrinsic to the screening method, and not necessarily shared by other methods (e.g. radiation-induced cancer due to radiographic screening). Interval cancers are not considered to be a harm, because they are, in essence, a planned outcome of the frequency with which screening is offered to members of

the target population and are balanced against harms that would increase in probability with increasing frequency of screening. However, it is recognized that some interval cancers are a consequence of a false-negative test.

The actual harms of the screening test itself or mediated by the screening-related events listed above include: (i) physical and psychological discomfort due to, and medical complications of, the screening method or further investigation of positive findings and subsequent treatment; (ii) all harmful consequences of overdiagnosis and/or overtreatment of screen-detected cancers, including preclinical cancers, and of precancerous lesions; (iii) unnecessary diagnosis and treatment of overdiagnosed cancers; and (iv) delay in diagnosis, a possibly poorer outcome of the targeted cancer, and feelings of betrayal due to the false reassurance of a false-negative finding.

Overdiagnosis is defined in the *Handbooks* as the diagnosis of a cancer as a result of screening that would never have caused any symptoms or problems if it had not been detected by screening. Screening may also detect a large number of precursors of cancer that would not have progressed to clinical cancer in the person's lifetime. The main concern in such cases is overtreatment. There are challenges to estimating overdiagnosis, and there are several ways in which it can be estimated, including the excess-incidence approach and the mean-lead-time approach. Estimates can be made from "well-conducted, population-based randomized controlled trials with long follow-up and minimal to no screening in the control group" (Davies et al., 2018), as well as from statistical modelling and from ecological studies. When there are several plausible estimates of overdiagnosis, results of any combined analyses of these estimates are also reviewed.

The IARC Secretariat, in consultation with the Working Group, may also commission or conduct a meta-analysis of such studies.

5.3 Balance of benefits and harms

A sound estimate of the balance of benefits and harms of a screening programme is important to aid decisions about whether to offer the programme and is most important for people who are deciding whether to participate in the programme. Estimates of the balance of benefits and harms for a particular cancer screening programme usually comprise one estimate of benefit (e.g. number of cancer deaths prevented per 1000 eligible people fully participating in the programme) and several estimates of harm (e.g. number of false-positive screening tests, and number of overdiagnosed cancers, per 1000 eligible people fully participating in the programme). These estimates are usually based on experimental or high-quality observational evaluations (e.g. incidence-based mortality analyses done under optimal circumstances) of the performance of screening methods or programmes. To project estimates of benefits and harms to a steady-state programme operating in a particular general population, modelling is required.

After identification of all published estimates of the balance of benefits and harms expressed in absolute terms (e.g. numbers of beneficial and harmful outcomes per 1000 screened individuals), the Working Group selects those based on the highest-quality evaluative studies of the commonly implemented screening regimens, critically assesses each study, summarizes the results in narrative or tabular format as appropriate, and critically assesses the body of evidence. The Working Group may also propose one or more "best" estimates of the balance of benefits and harms, while noting the limits of applicability of those estimates to settings other than the populations and screening experience from which they were derived.

As noted in Part B, Section 1, the balance of benefits and harms of screening is expected to be more favourable in organized screening programmes than in the case of opportunistic screening. The balance may also differ substantially between specific population subgroups, for example human papillomavirus (HPV)-vaccinated and non-vaccinated women for cervical cancer screening. Major factors that influence the balance of benefits and harms include background cancer risk, life expectancy, sex, and age. Where possible, the Working Group will acknowledge these factors and consider comparing benefits and harms for different population subgroups.

In addition to the balance of benefits and harms, the net benefit of screening (which can be positive or negative) may be estimated in an aggregate manner, for example by calculating the average number of

quality-adjusted life years (QALYs) gained or disability-adjusted life years (DALYs) averted as a result of screening. QALYs and DALYs are generic measure of disease burden that include quality and quantity of life in their estimation. Because both are based on estimation of lifetime outcomes and are estimated by modelling, they cannot be estimated directly from trials.

In consultation with the Working Group and when it is feasible and potentially contributory, the IARC Secretariat may commission or conduct a systematic review of modelling studies that have estimated QALYs gained or DALYs averted from screening, and also modelling studies that have estimated disaggregated measures of benefits or harms. The Working Group will critically appraise the quality of the studies using internationally accepted criteria for good modelling conduct (Caro et al., 2012) and applicable subject-specific quality frameworks for models. High-quality collaborative modelling studies (i.e. studies in which different modelling groups work together using standardized assumptions) will be favourably viewed in considering the overall quality of a particular evaluation. Petitti et al. (2018) provided a checklist for the critical appraisal of collaborative modelling reports specific to cancer screening, which can also be used for the appraisal of single modelling studies. Baseline parameters used and their sources, most particularly the sources of calibration data, and other assumptions made in the absence of relevant baseline data require careful scrutiny. Special attention needs to be paid to the extent to which weights for quality and disability have been incorporated for all relevant phases of screening and management of cancer, and also whether disutility is available for all downstream management pathways after the screening test, and whether these have been modelled in detail or as a single aggregate disutility. Currently, there is a general paucity of evidence to support detailed modelling of disutility for each step involved in screening, triage, diagnosis, surveillance, and treatment (all of which are required to model the detailed impact of a screening programme on QALYs or DALYs). As a result, primary studies may judiciously choose to present aggregate benefits information summarized as life years saved, and these data should be considered very carefully as less prone to issues around the uncertainty inherent in estimation of QALYs or DALYs.

5.4 Cost–effectiveness

For a screening method or programme that is capable of delivering a beneficial outcome, cost–effectiveness is usually expressed as the estimated financial cost of implementing the method or programme per unit of the benefit it delivers, which is most often measured in terms of life years, as QALYs gained or DALYs averted. The ratio of costs to benefits (i.e. level of cost–effectiveness) needed to implement a health service programme varies from country to country, depending principally on the wealth of the country and on who pays (e.g. the government or individual citizens). Therefore, the specific ratio derived from cost–effectiveness analyses from a certain country is usually not generalizable to other countries and settings. However, if there are sufficient (high-quality) analyses from different parts of the world with consistent results on the cost–effectiveness of the screening intervention of interest within their respective settings, qualitative statements can be made about the cost–effectiveness of the screening intervention. Although assessments of cost–effectiveness that account for all costs (e.g. that are not restricted to health service costs) are less frequently done, it is important to note that their perspective may differ markedly from one based on health service costs only. Like the balance of benefits and harms, cost–effectiveness estimates can be markedly different in different population subgroups, depending on background cancer risk, life expectancy, sex, and age, among others. Ideally, the cost–effectiveness analysis should be based on the primary population targeted for screening; incremental analyses can consider the inclusion of additional populations (e.g. extended age range for screening).

Taking a similar approach to that taken for the balance of benefits and harms described above, the IARC Secretariat may commission or conduct a systematic review of published reports of cost–effectiveness analyses. Studies to be included report on net costs (including upfront costs of screening and downstream costs and savings for follow-up and management of cancers) as well as net benefits, preferably in the form of life years gained, QALYs, or DALYs. Methods for all such studies will include modelling. Where applicable, study quality will be appraised in ways similar to those described in Section 5.1b, with the addition of appraisal against internationally accepted criteria for good conduct of cost–effectiveness analysis, such as the Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses by the Second Panel on Cost-Effectiveness in Health and

Medicine (Sanders et al., 2016). Methods, assessment against quality criteria, and results will be tabulated for high-quality studies of commonly implemented screening regimens. To ensure sufficient regional variation in the reports, low-quality cost–effectiveness analyses may also be reported and considered in the overall assessment of cost–effectiveness for regions without high-quality reports. The results do not contribute to the overall evaluation of each screening method but can be used by governments and health services to aid decisions about implementation of screening for which there is sufficient evidence of a screening effect.

5.5 Comparison of effects of separately reviewed screening methods

When two screening methods have been established to reduce cancer incidence or cancer mortality, an evaluation may be conducted of the comparative efficacy or effectiveness of these methods. Studies that compare the effects of screening of two or more different screening methods are reviewed and rigorously assessed. Where possible, a statement is made as to the strength of the evidence that use of one screening method is more efficacious or effective than use of another, together with an evaluation of any comparative data about additional dimensions, such as screening protocol, acceptability, harms, costs, and equity of access, that can influence the population impact of a screening method.

In the absence of such evidence, the Working Group may critically appraise the commonly advanced reasons for choosing one method over another and the justifications given for them, taking into account all the dimensions listed above.

5.6 Surveillance in populations at increased risk

Screening in people with a personal history of the cancer type subject to screening is not evaluated in the *Handbooks*.

Population subgroups at substantially increased risk of the target cancer(s) are briefly described. Available evidence relating to the effect of screening in any of these populations using any of the separately considered screening methods is systematically reviewed and analysed with the same rigour as evidence in whole populations or populations at average risk, and, where possible, a statement is made as to the strength of the evidence that use of any screening method or particular screening method regimen in the group at high risk is more efficacious or effective than use of any other screening method or regimen. Where possible, the magnitudes of the benefits and the harms of the screening method or regimen in these populations are given.

In the absence of such evidence, the Working Group may critically appraise approaches commonly taken to screening in defined groups at high risk and the justifications that have been given for them.

5.7 Other topics reviewed

Some other topics important to the practice of screening may be reviewed in a *Handbook* by summarizing a representative set of studies. These topics do not contribute to the overall evaluations of the screening methods. They may include, among others:

(a) Determinants of participation in screening

Given an often large and complex literature, a review of reviews of studies in high-income populations and of individual studies from low- and middle-income countries is performed. Special attention is given to the impact on equity of access to effective screening when assessing the role of barriers and the effectiveness of interventions aimed at promoting participation.

(b) Quality of life

The results of studies on gain or loss in quality of life of participants in screening programmes that add useful information on the value of screening are reviewed. Only a few studies have directly investigated change in quality of life as an outcome of screening programmes. These estimates can be used in health (economic) assessments as disability weights when estimating DALYs, QALYs, and cost–effectiveness. Although the available quality-of-life studies usually address physical, social, and emotional functional abilities and general satisfaction, the assessment of health-related quality of life gained or lost through screening programmes is challenging and is heavily context-dependent.

6. Summary of data reported

Each section or subsection of the *Handbook* is summarized. The cancer type subject to screening and its global burden are described, the screening methods evaluated are identified, and their global use is briefly presented. The results of epidemiological studies addressing the efficacy, effectiveness, and harms of each screening method are also summarized. The overall strengths and limitations of the epidemiological evidence base are highlighted to indicate how the evaluation was reached. Typically, the relative and absolute reductions in incidence and/or mortality in populations adhering to the screening regimen evaluated are presented. Harms of the screening intervention are described, both qualitatively and quantitatively, as the evidence base permits.

Depending on the amount and relevance of the data, the Working Group may also summarize the reviewed evidence for cost–effectiveness, and for any other item that the Working Group considers sufficiently important to note.

7. Evaluation and rationale

Although the following details about the evaluation and rationale refer specifically to screening interventions, they will also apply for the evaluation of early diagnosis interventions, with some adaptation as needed.

Consensus evaluations of the strength of the evidence of a reduction of cancer incidence and/or cancer mortality (preventive effects) in humans of each screening method reviewed are made using transparent criteria and defined descriptive terms (see below). Statements should also be made about the evidence for harms and for the balance of benefits and harms.

Where the evaluation of several cancer screening methods indicates that they can reduce cancer incidence and/or cancer mortality (Group A; see below), the Working Group may also choose to indicate whether the efficacy or effectiveness in reducing cancer incidence and/or cancer mortality and the balance of benefits and harms of one screening method are superior to those of another screening method.

Similarly, the Working Group may choose to evaluate the efficacy or effectiveness of one screening method or protocol implemented in a population at increased risk of the cancer, depending on whether relevant evidence is available.

The framework for these evaluations, described below, may not encompass all factors relevant to a particular evaluation of preventive efficacy or effectiveness. After considering all relevant scientific findings, the Working Group may exceptionally assign the intervention to a different category than a strict application of the framework would indicate, while providing a clear rationale for such an evaluation.

The wording of these evaluations is the same when inferences about preventive effects are made from the results of studies in which an intermediate outcome, not cancer incidence and/or cancer mortality, was the outcome studied. Such evaluations are made only when a causal association has been established between the intermediate outcome and cancer. A statement to this effect is added.

The evaluation is followed by a description or discussion of harms, with a qualitative and quantitative overall evaluation considered in the light of potential and actual harms.

When there are substantial differences of scientific interpretation among the Working Group members, the overall evaluation will be based on the consensus of the Working Group. A summary of the alternative interpretations may be provided, together with their scientific rationale and an indication of the degree of support for each.

The evaluation categories refer to the strength of the evidence that an intervention can reduce the incidence of cancer or death from cancer; they do not address how strongly or weakly the intervention reduces cancer incidence and/or cancer mortality, if it can. Put another way, they do not address the question "By how much might or does this intervention reduce cancer incidence or cancer mortality in exposed people?"

7.1 Evaluation

On the basis of the principles outlined in Part B, Section 5, the evidence relevant to cancer prevention is classified into one of the following categories:

(i) The cancer screening method *is established to reduce the incidence of cancer of the [target organ] OR is established to reduce mortality from cancer of the [target organ]* **(Group A)**

A causal preventive association between use of the screening method or screening methods and cancer incidence or mortality has been established. That is, a preventive association has been observed consistently in the body of evidence on use of the screening method or methods and cancer incidence or mortality, and chance, bias, and confounding as explanations for the association were ruled out with reasonable confidence.

When the evidence is classified in **Group A**, the evaluation is followed by separate sentences to:

• make a statement as to the screening regimen to which the Working Group considers each evaluation of a screening method applies or applies most strongly, and as to whether or not the effectiveness of that screening method has been established;

• make a statement of what the Working Group considers to be the magnitudes of the benefits and the harms of the screening method, in as nearly comparable terms as possible, for people adhering fully to the screening approach most commonly implemented in practice, and whether or not the benefits outweigh the harms.

(ii) The cancer screening method *may reduce the incidence of cancer of the [target organ] OR may reduce mortality from cancer of the [target organ]* **(Group B)**

A causal preventive association between use of the screening method or methods and cancer incidence or mortality is credible, but chance, bias, or confounding as explanations for the association could not be ruled out with reasonable confidence; OR a causal preventive association between use of the screening method and incidence of precancer or clinically advanced cancer has been established in the absence of an established association for cancer incidence or mortality, respectively.

When the evidence is classified in **Group B**, a sentence makes a statement as to the screening regimen to which the Working Group considers each evaluation of a screening method (or of closely related methods collectively, when evaluated together) applies or applies most strongly.

(iii) The cancer screening method *is not classifiable as to its capacity to reduce the incidence of cancer of the [target organ] OR to reduce mortality from cancer of the [target organ]* **(Group C)**

The available studies are of insufficient quality, consistency, or statistical precision to enable a conclusion to be drawn about the presence or absence of a causal preventive association between the screening method or methods and cancer incidence or mortality; OR there is some evidence that the screening method or methods has a preventive effect, based on precancer or clinically advanced cancer as outcomes, but not enough to qualify for Group B. The first of the above conditions includes: (a) there are relevant studies available, but all are of poor quality or informativeness; and (b) there are relevant studies available of sufficient quality, but their results are inconsistent or otherwise inconclusive.

(iv) The cancer screening method *may lack the capacity to reduce the incidence of cancer of the [target organ] OR to reduce mortality from cancer of the [target organ]* **(Group D)**

There are several high-quality studies that are mutually consistent in not showing a preventive association between the screening method or methods and the studied cancer at the observed levels of use. The results from these studies alone or combined should have narrow confidence intervals with upper limits above or close to the null value (e.g. a relative risk of 1.0). Chance, bias, and confounding as explanations for the null results were ruled out with reasonable confidence, and the studies were considered informative. Consistent and substantial evidence that the screening method does not result in diagnosis that is earlier in the natural history of cancer than is observed in the absence of screening OR that

cancer-specific survival of cancers detected by screening is no better than that of cancers diagnosed in the absence of screening also provide evidence for lack of cancer prevention from the screening method.

A conclusion that the screening method *may lack the capacity to reduce cancer incidence and/or cancer mortality* is limited to the screening method or methods evaluated and the populations and life-stages, conditions and levels of screening, and length of observation covered by the available studies. In addition, the possibility of a very small preventive effect at the levels of the intervention studied can never be excluded.

7.2 Rationale

The reasoning that the Working Group uses to reach its evaluation is summarized so that the basis for the evaluation offered is transparent. This section includes concise statements of the principal lines of argument that emerged in the deliberations of the Working Group, the conclusions of the Working Group on the strength of the evidence, an indication of the body of evidence that was pivotal to these conclusions, and an explanation of the reasoning of the Working Group in making the evaluations. Where relevant, it also includes reference to use of an intermediate outcome as an, or the, evaluation outcome.

In the rationale, the Working Group may draw attention to the fact that the evaluations should be interpreted in the light of specific circumstances that vary between countries, which influence the feasibility of implementation of programmes based on the interventions evaluated.

References

AMSTAR (2017). AMSTAR checklist. Assessing the methodological quality of systematic reviews. Available from: https://amstar.ca/Amstar_Checklist.php, accessed 6 March 2019.

Axelson O, Steenland K (1988). Indirect methods of assessing the effects of tobacco use in occupational studies. Am J Ind Med. 13(1):105–18. https://doi.org/10.1002/ajim.4700130107 PMID:3344750

Boyle P, Levin B, editors (2008). World cancer report 2008. Lyon, France: International Agency for Research on Cancer. Available from: http://publications.iarc.fr/395.

Caro JJ, Briggs AH, Siebert U, Kuntz KM; ISPOR-SMDM Modeling Good Research Practices Task Force (2012). Modeling good research practices – overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. Med Decis Making. 32(5):667–77. https://doi.org/10.1177/0272989X12454577 PMID:22990082

Cochrane AL (1972). Effectiveness and efficiency: random reflections on health services. Nuffield Trust. Available from: https://www.nuffieldtrust.org.uk/research/effectiveness-and-efficiency-random-reflections-on-health-services.

Cogliano V, Baan R, Straif K, Grosse Y, Secretan B, El Ghissassi F, et al. (2005). Transparency in IARC Monographs. Lancet Oncol. 6(10):747. https://doi.org/10.1016/S1470-2045(05)70380-6

Cogliano VJ, Baan RA, Straif K, Grosse Y, Secretan MB, El Ghissassi F, et al. (2004). The science and practice of carcinogen identification and evaluation. Environ Health Perspect. 112(13):1269–74. https://doi.org/10.1289/ehp.6950 PMID:15345338

Cole P, Morrison AS (1980). Basic issues in population screening for cancer. J Natl Cancer Inst. 64(5):1263–72. PMID:6767876

Davies L, Petitti DB, Martin L, Woo M, Lin JS (2018). Defining, estimating, and communicating overdiagnosis in cancer screening. Ann Intern Med. 169(1):36–43. https://doi.org/10.7326/M18-0694 PMID:29946705

Emery JD, Gray V, Walter FM, Cheetham S, Croager EJ, Slevin T, et al. (2014). The Improving Rural Cancer Outcomes (IRCO) Trial: a factorial cluster-randomised controlled trial of a complex intervention to reduce time to diagnosis in rural patients with cancer in Western Australia: a study protocol. BMJ Open. 4(9):e006156. https://doi.org/10.1136/bmjopen-2014-006156 PMID:25231496

Greenland S, O'Rourke K (2008). Meta-analysis. In: Rothman KJ, Greenland S, editors. Modern epidemiology. Philadelphia (PA), USA: Wolters Kluwer Health/Lippincott Williams & Wilkins; pp. 652–82.

Hernán MA, Hernández-Díaz S, Robins JM (2004). A structural approach to selection bias. Epidemiology. 15(5):615–25. https://doi.org/10.1097/01.ede.0000135174.63482.43 PMID:15308962

Hill AB (1965). The environment and disease: association or causation? Proc R Soc Med. 58:295–300. PMID:14283879

IARC (1997). Non-steroidal anti-inflammatory drugs. IARC Handb Cancer Prev. 1:1–202. Available from: http://publications.iarc.fr/371.

IARC (2002a). Weight control and physical activity. IARC Handb Cancer Prev. 6:1–315. Available from: http://publications.iarc.fr/376.

IARC (2002b). Breast cancer screening. IARC Handb Cancer Prev. 7:1–229. Available from: http://publications.iarc.fr/377.

IARC (2005). Use of screening for cervical cancer. In: Cervix cancer screening. IARC Handb Cancer Prev. 10:117–62. Available from: http://publications.iarc.fr/380.

IARC (2007). Tobacco control: reversal of risk after quitting smoking. IARC Handb Cancer Prev. 11:1–358. Available from: http://publications.iarc.fr/381.

IARC (2008). Methods for evaluating tobacco control policies. IARC Handb Cancer Prev. 12:1–300. Available from: http://publications.iarc.fr/382.

IARC (2009). Evaluating the effectiveness of smoke-free policies. IARC Handb Cancer Prev. 13:1–334. Available from: http://publications.iarc.fr/383.

IARC (2011). Effectiveness of tax and price policies for tobacco control. IARC Handb Cancer Prev. 14:1–366. Available from: http://publications.iarc.fr/384.

IARC (2016a). Breast cancer screening. IARC Handb Cancer Prev. 15:1–469. Available from: http://publications.iarc.fr/545.

IARC (2016b). Study designs to assess the effectiveness of screening. In: Breast cancer screening. IARC Handb Cancer Prev. 15:287–94. Available from: http://publications.iarc.fr/545.

IARC (2018). Guidelines for Observers at IARC Handbooks Meetings. Available from: http://handbooks.iarc.fr/workingprocedures/guidelines_observers.php.

IARC (2019a). International Classification of Diseases for Oncology, ICD-O-3 online. Lyon, France: International Agency for Research on Cancer. Available from: http://codes.iarc.fr/usingicdo.php, accessed 8 January 2019.

IARC (2019b). WHO Classification of Tumours. Lyon, France: International Agency for Research on Cancer. Available from: http://whobluebooks.iarc.fr/, accessed 8 January 2019.

Irwig L, Houssami N, Armstrong B, Glasziou P (2006). Evaluating new screening tests for breast cancer. BMJ. 332(7543):678–9. https://doi.org/10.1136/bmj.332.7543.678 PMID:16565097

Miles A, Cockburn J, Smith RA, Wardle J (2004). A perspective from countries using organized screening programs. Cancer. 101(5 Suppl):1201–13. https://doi.org/10.1002/cncr.20505 PMID:15316915

Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097 PMID:19621072

Petitti DB, Lin JS, Owens DK, Croswell JM, Feuer EJ (2018). Collaborative modeling: experience of the U.S. Preventive Services Task Force. Am J Prev Med. 54(1S1):S53–62. https://doi.org/10.1016/j.amepre.2017.07.003 PMID:29254526

Porta M, editor (2008). A dictionary of epidemiology. 5th ed. New York (NY), USA: Oxford University Press. Available from: http://www.oxfordreference.com/view/10.1093/acref/9780195314496.001.0001/acref-9780195314496.

Porta M, editor (2014). A dictionary of epidemiology. 6th ed. New York (NY), USA: Oxford University Press. Available from: https://global.oup.com/academic/product/a-dictionary-of-epidemiology-9780199976737?cc=fr&lang=en&.

Richardson DB, Laurier D, Schubauer-Berigan MK, Tchetgen Tchetgen E, Cole SR (2014). Assessment and indirect adjustment for confounding by smoking in cohort studies using relative hazards models. Am J Epidemiol. 180(9):933–40. https://doi.org/10.1093/aje/kwu211 PMID:25245043

Rothman KJ, Greenland S, Lash TL (2008). Causation and causal inference. In: Modern epidemiology. Philadelphia (PA), USA: Wolters Kluwer Health/Lippincott Williams & Wilkins; pp. 5–31.

Sanders GD, Neumann PJ, Basu A, Brock DW, Feeny D, Krahn M, et al. (2016). Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. JAMA. 316(10):1093–103. https://doi.org/10.1001/jama.2016.12195 PMID:27623463

Siemieniuk R, Guyatt G (2019). What is GRADE? BMJ Best Practice. Available from: https://bestpractice.bmj.com/info/toolkit/learn-ebm/what-is-grade/, accessed 6 March 2019.

Stewart BW, Kleihues P, editors (2003). World cancer report 2003. Lyon, France: International Agency for Research on Cancer. Available from: http://publications.iarc.fr/397.

Stewart BW, Wild CP, editors (2014). World cancer report 2014. Lyon, France: International Agency for Research on Cancer. Available from: http://publications.iarc.fr/396.

Vandenbroucke JP, Broadbent A, Pearce N (2016). Causality and causal inference in epidemiology: the need for a pluralistic approach. Int J Epidemiol. 45(6):1776–86. https://doi.org/10.1093/ije/dyv341 PMID:26800751

WHO (2017). Guide to cancer early diagnosis. Geneva, Switzerland: World Health Organization. Available from: https://www.who.int/cancer/publications/cancer_early_diagnosis/en/

Wild CP, Cogliano VJ (2011). A plea on behalf of expert evaluation and the experts involved. Int J Epidemiol. 40(1):253–61. https://doi.org/10.1093/ije/dyq038 PMID:20223796

Young GP, Senore C, Mandel JS, Allison JE, Atkin WS, Benamouzig R, et al. (2016). Recommendations for a step-wise comparative approach to the evaluation of new screening tests for colorectal cancer. Cancer. 122(6):826–39. https://doi.org/10.1002/cncr.29865 PMID:26828588