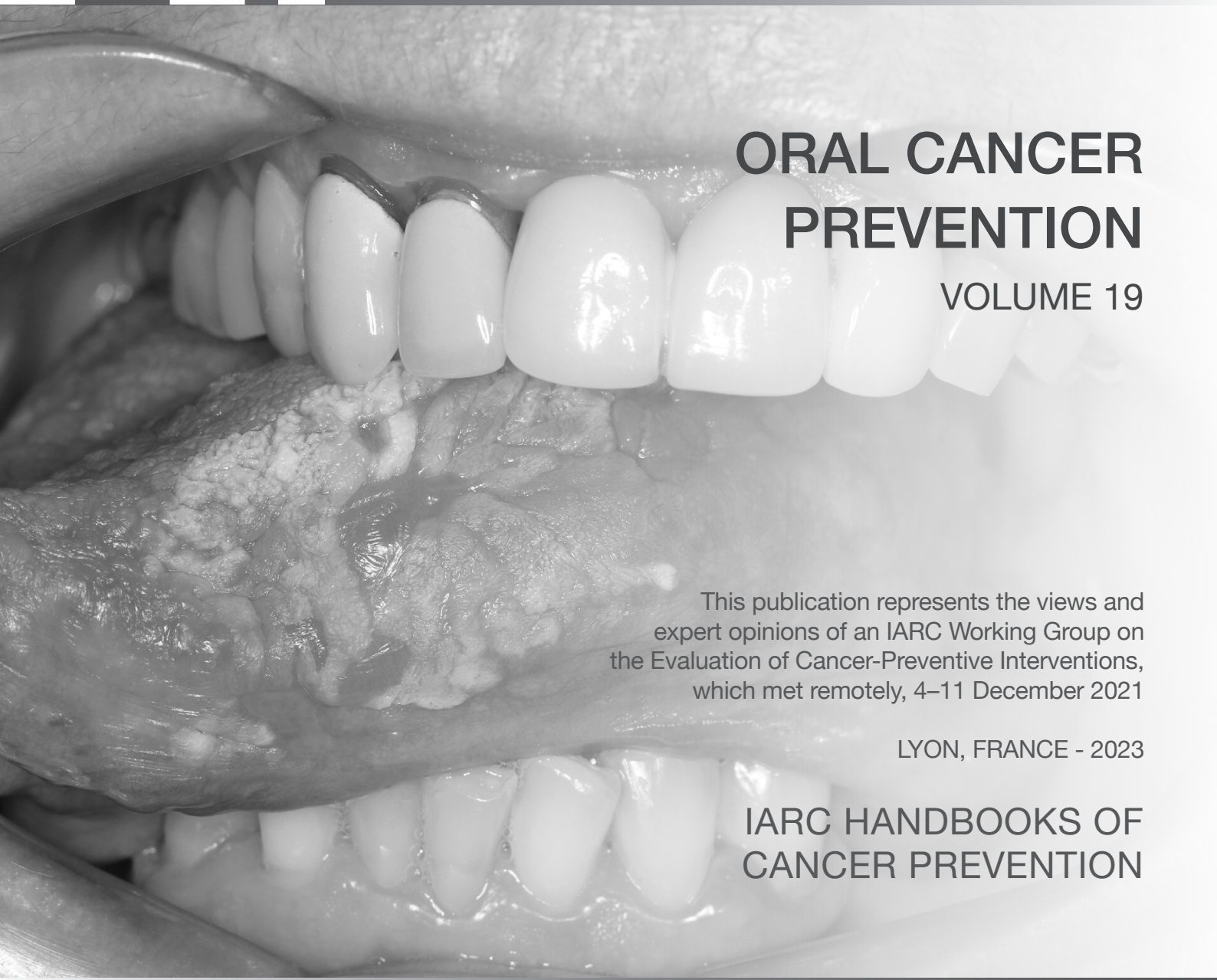
IARC HANDBOOKS



ORAL CANCER PREVENTION

VOLUME 19

This publication represents the views and expert opinions of an IARC Working Group on the Evaluation of Cancer-Preventive Interventions, which met remotely, 4–11 December 2021

LYON, FRANCE - 2023

IARC HANDBOOKS OF CANCER PREVENTION

International Agency for Research on Cancer

World Health Organization

PREAMBLE – PRIMARY 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, 2016), 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 be sufficiently robust to encompass a wide variety of interventions, ranging from broad societal measures to individual behaviour and to chemoprevention.

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, understanding important 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 primary prevention

The current *IARC Handbook* addresses a specific intervention or class of interventions for **primary prevention**. Primary prevention "aims to reduce the incidence of disease by personal and communal efforts" (Porta, 2014). The term "intervention" in this *Handbook* refers to any action aimed at reducing the incidence of cancer in humans. Primary prevention interventions include increasing human exposure to known cancer-preventive agents, reducing human exposure to known cancer hazards, providing means to reduce the effects of exposure to cancer hazards, or otherwise intervening on human pathological states that cause cancer. In broad terms, such interventions include, for example, regulating exposure to carcinogens, administering chemopreventive pharmaceuticals or other agents, vaccinating against cancer-causing infections, modifying the environment (e.g. planting trees or constructing shade structures in areas of high ambient levels of solar ultraviolet radiation), or promoting personal or societal action to increase the prevalence of healthy lifestyles or behaviours or decrease the prevalence of unhealthy lifestyles or behaviours.

Primary preventive 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.

Although the intent of the *IARC Handbooks* is to evaluate interventions, i.e. a dynamic comparison, there will be circumstances under which an evaluation of the association between exposure to an agent and cancer incidence, i.e. a static comparison, is appropriate. In principle, the approaches to scientific review of the relevant studies in this section will not differ between those entailing dynamic interventions and those entailing static exposures. Therefore, in this Preamble the term "intervention" applies to studies of both types, unless specifically stated otherwise.

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 recommending 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, preventive 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.

Other issues to be considered include the cost, cost–effectiveness, affordability, economic efficiency, health equity impact, feasibility, acceptability, relative value, and human rights impact of the intervention. Depending on the specific intervention, some of these issues may be of sufficiently high interest to be reviewed in the *IARC Handbook*.

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 process of the *IARC Handbooks*, 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, IARC defines the intervention and the outcome to be evaluated, according to one of two scenarios:

Scenario 1: evaluation of the effect of a specified *intervention*, that is, an action that results in a change in a potentially preventive exposure, in producing a specified change in *cancer incidence*.

Scenario 2: a two-step evaluative framework from which, for scientific reasons, the level of evidence that an intervention prevents cancer is established by way of an intermediate outcome.

- In Step 1, the effect of a specified intervention on an intermediate outcome, such as exposure to a particular risk factor or preventive factor for cancer in humans, is evaluated (Jonas et al., 2018). Step 1 alone might be taken if it **has been established in authoritative sources** (e.g. the *IARC Monographs* programme) that a change in the intermediate outcome (decreasing exposure to a risk factor or increasing exposure to a preventive factor) reduces the risk of cancer in humans.
- In Step 2, the effect of the change in the intermediate outcome (decrease in exposure to the risk factor or increase in exposure to the preventive factor) on cancer incidence in humans is evaluated. Evaluation of data streams to support Step 2 alone might be done in preparation for a subsequent evaluation of data to support Step 1 if it **has not yet been established in authoritative sources** that a

change in the intermediate outcome reduces the risk of cancer in humans.

The analytical framework determines whether evidence is reviewed for Step 1 only, Step 2 only, or both Steps 1 and 2. A *Handbook* might, for example, include both Steps 1 and 2 when a systematic review and evaluation of Step 2 is necessary (e.g. is not yet available from other authoritative sources) and the number of studies to be reviewed for Steps 1 and 2 is manageable. Taking Steps 1 and 2 together is equivalent to Scenario 1 with inclusion of one or more intermediate outcomes in the evaluation scheme. The sections below provide additional details on the selection of the interventions and outcomes for review.

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). IARC will also pay attention to topics that extend beyond those covered by other agencies.

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 have not been established formally;
- The available evidence suggests that the intervention has the potential to significantly reduce the incidence of cancer, or to have a significant impact on an intermediate outcome or outcomes known or highly suspected to be linked to cancer (see Section 3.1; see also 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 one or several specific cancer sites or to specific aspects of the preventive intervention (e.g. reduction in excess body fatness) to which the new evidence predominantly relates. 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.

3.3 Selection of the outcomes

In primary prevention of cancer, the outcome targeted by the preventive intervention or interventions is reduction in the incidence of cancer (Scenario 1; see Part A, Section 3.1).

As described above, an intermediate outcome may be chosen as the evaluation outcome for a *Handbook* when there is evidence that a change in the intermediate outcome (decreasing exposure to the risk factor or increasing exposure to the preventive factor) can lead to a reduction in the incidence of one or more types of cancer. An example of such a target is an increase in the smoking cessation rate, which is a commonly used outcome for studies designed to determine the preventive effects of new methods of reducing the incidence of tobacco-caused cancer

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

by way of reducing the prevalence of tobacco smoking. Other examples of changes in intermediate outcomes include a decrease in excess body fatness, a decrease in the levels of diesel engine emissions in urban environments, and an increase in the population coverage of human papillomavirus (HPV) vaccination.

Alternatively, a *Handbook* could, as a first step, evaluate the evidence that changing the intermediate outcome can lead to a reduction in the incidence of one or more types of cancer if such evidence is not already available from authoritative sources, followed by an evaluation of the effect of an intervention on the intermediate outcome (Scenario 2, Step 2 followed by Step 1; see Part A, Section 3.1). An example of such a scenario is evaluation of the evidence that reducing consumption of alcoholic beverages reduces incidence of alcohol-related cancer or precancer, followed by evaluation of the efficacy or effectiveness of a specific intervention in reducing the consumption of alcoholic beverages.

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). 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 or affect; 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 interest 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.

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 (https://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 interest, for transparency and to provide an opportunity for undeclared conflicts of interest 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

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

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.

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 2–6 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 pertinent reports and reviews on: use of the intervention targeted directly to cancer or to a relevant intermediate outcome or outcomes; all experimental and observational studies in humans (including systematic reviews and meta-analyses) of the putative effect of the intervention or interventions on cancer incidence or a relevant intermediate outcome, and any related harms; all relevant experimental studies in animals; and all relevant mechanistic studies.

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, and encompasses the inclusion of studies in humans, studies in experimental animals, and mechanistic studies when relevant, 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 pharmaceuticals), 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 2–4. Study design and results are tabulated systematically in a standard format. This step may be iterative with Step 5.

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 2–4. 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 2.1d.

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 is described in terms given in Part B, Sections 6a–c for each stream of evidence. The Working Group then integrates the strength-of-evidence conclusions from all streams of evidence (see Part B, Section 6d) and develops the rationale for its overall consensus evaluation of the cancer-preventive effect of the intervention (see Part B, Sections 6d–e).

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 5), using the *IARC Handbooks* criteria (see Part B, Section 6a–c).

(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, Sections 6a–c), 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 6d.

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. Intervention and outcome characterization

An intervention for primary cancer prevention has been defined in this Preamble to be any action aimed at reducing the incidence of cancer in humans (Part A, Section 2). Given this definition, the efficacy or effectiveness of an intervention would be most directly approached by research that examines whether the delivery of the intervention results in a measurable change in a cancer-related exposure that leads to a reduction in the incidence of cancer. However, such research is often lacking, and therefore the possibility of cancer-preventive effects has often been inferred from static associations of cancer incidence with prevalence of exposure to cancer-causing agents or cancer-preventive agents. For example, all measures that are now taken to minimize environmental exposure to asbestos (e.g. regulation of removal of asbestos from buildings or demolition of buildings known to contain asbestos) are based on the very strong evidence that people who have had identifiable exposure to asbestos have a higher incidence of cancer than people who have not had such exposure. Similarly, the evaluation of *Handbooks* Volume 16 that there "is *sufficient evidence* in humans for a cancer-preventive effect of absence of excess body fatness" is almost exclusively based on the substantial body of evidence that cancer incidence is lower in people without excess body fatness than it is in people with excess body fatness; this is a static comparison, not a dynamic comparison as the term "intervention" implies.

1.1 Intervention characterization

This section provides informative background on the intervention and the factors that mediate it. It also summarizes the prevalence and level of the intervention across geographical areas and across the life-course. Methods used to assess exposure to the intervention in key experimental and observational epidemiological studies are described and evaluated. This section also reports on validated biomarkers of internal exposure, metabolites, or other intermediate outcomes that are routinely used for exposure assessment. Concepts of absorption, distribution, metabolism, and excretion, where relevant, are considered in the section on mechanistic evidence (see Part B, Section 4b).

(a) Identification of the intervention

The intervention being evaluated is unambiguously identified. The information provided will vary widely depending on the type of intervention but should be sufficient to enable the implementation of an intervention in practice with reasonable confidence that its outcomes in populations would be similar to those of the intervention from which the bulk of the evidence evaluated in the *Handbook* originated.

Many interventions are multifaceted and comprise complex sets of actions. Interventions determined by personal behaviour or circumstances may result from, be influenced by, or be correlated with a diverse range of behavioural and environmental factors, such as smoking, alcohol consumption, diet, sleep and physical activity patterns, remoteness of residence, and socioeconomic circumstances. The description of such interventions should include their variability across human populations and environments, and their known relationships with other health-determining factors.

(b) Global occurrence and use

Geographical patterns and time trends in occurrence are summarized. A concise overview of quantitative information about sources, prevalence, and levels of individual and population interventions, whether purposive or incidental, is provided. Representative data from formal environmental or behavioural monitoring or surveillance data, research studies, government reports and websites, online databases, and other

citable, publicly available sources are tabulated. Data from low- and middle-income countries are sought and included to the extent that is feasible; information gaps for key regions are noted.

If available, data are reported by region and by other relevant characteristics, such as sex, age, socioeconomic status, and other variables considered relevant by the Working Group.

(c) Regulations and guidelines

Regulations or guidelines that have been established for the intervention (e.g. permissible levels of fortification in food, national dietary guidelines) are described and may be tabulated if they are informative for the interpretation of current or historical levels of the intervention. Information on applicable populations, the basis for regulation, and the timing of regulation may be noted.

(d) Intervention assessment in key epidemiological studies

Epidemiological studies reviewed in the context of the *IARC Handbooks* programme evaluate cancer prevention interventions (or effects on intermediate outcomes) by comparing outcomes across groups differently exposed to changes in a putative cancer-preventing intervention. Therefore, the type and the quality of intervention assessment methods used are key considerations when interpreting study findings. This section summarizes and critically reviews the intervention assessment methods used in both experimental and observational epidemiological studies that contribute data relevant to the *Handbooks* evaluation.

All interventions have two principal dimensions: (i) dose (sometimes defined as concentration or intensity), and (ii) time considerations, including duration (time from first to last exposure), pattern or frequency (whether continuous or intermittent), and windows of susceptibility. This section considers how each of the key epidemiological studies characterizes these dimensions. Interpretation of information for chemical, biological, or physical interventions may also be informed by consideration of mechanistic evidence on absorption, distribution, metabolism, and excretion (e.g. as described in Part B, Section 4b).

In experimental epidemiological studies, the investigators determine, usually by way of randomization, who will and who will not be assigned to the intervention; however, in practice the assignment is not always adhered to. Therefore, a critical assessment of such studies requires careful evaluation using appropriate guidelines or assessment frameworks (e.g. fidelity to intervention implementation and extent of non-adherence to intervention).

Intervention intensity and timing in observational epidemiological studies can be characterized by using environmental monitoring data, records from workplaces or other sources, and subject or proxy reports collected by way of questionnaires or interviews. Both objective and subjective data sources are used, individually or in combination, to assign levels or values of an intervention metric to members of the study population.

Key epidemiological studies with interventions on cancer or intermediate outcomes are identified, and the intervention assessment approach and its strengths and limitations are summarized in text and tables. The Working Group identifies concerns about intervention assessment methods and their impacts on the overall quality of each study reviewed. The Working Group notes the studies where the information provided to characterize the intervention properly, the adherence to the intended intervention in each arm of experimental studies, or the assessment of the intervention in observational studies is inadequate. The Working Group further discusses the likely direction of bias due to non-adherence or to error in intervention assessment in studies where adequate information is available.

1.2 Outcome characterization

(a) Evaluation of cancer outcomes

The cancers are defined and described in terms of their International Classification of Diseases for Oncology (ICD-O) (IARC, 2019) or International Classification of Diseases (ICD) categories, with other relevant morphological or molecular characteristics where relevant.

Benign neoplasms, pre-neoplastic lesions, malignant precursors, and other end-points closely related to cancer may also be reviewed when they relate to the intervention reviewed and are known to predict the primary cancer outcome. These studies can strengthen evidence from studies of cancer itself. For example, the results of controlled trials of sun protection measures in preventing development of cutaneous melanocytic naevi (which are strong risk factors for development of later cutaneous melanoma) in children provide support for the efficacy of sun protection measures in preventing cutaneous melanoma in adults (Thun et al., 2018).

(b) Evaluation of intermediate outcomes

Potentially relevant intermediate outcomes vary widely across human biology, pathology, and behaviour. (Intermediate outcomes that are biomarkers of early biological effects, which are not topics evaluated in *IARC Handbooks*, are described in Part B, Section 4.) All intermediate outcomes are described as precisely as possible, using an applicable international standard classification (e.g. ICD classification). When, as with some behavioural or physiological risk factors, they can be defined or measured in a range of ways, the definitions that are acceptable for the evaluation are clearly defined and acceptable standards for measurement stated.

When an intermediate outcome is the outcome being evaluated, the evidence base establishing that the intermediate outcome has an established causal or preventive association with cancer incidence is briefly summarized.

In what follows, the term "cancer incidence" refers to the **outcome of a *Handbooks evaluation***, that is, to the incidence of **cancer** or of an **intermediate outcome**, as defined in the analytical framework.

2. Studies of cancer prevention in humans

This section includes all pertinent experimental and observational studies in humans that include cancer or a specified intermediate outcome (if it is the topic of the *Handbook*) as a study outcome. As noted above, only observational studies in which changes in the exposure (i.e. intervention) in relation to the outcome have been analysed will be considered, unless specifically stated otherwise. Among many others, these studies also encompass studies with biomarkers as intervention metrics (Alexandrov et al., 2016). As mentioned above, studies that assess biomarkers of early biological effects are reviewed in Part B, Section 4.

This section includes specification and assessment of beneficial effects, as well as potential harms.

2.1 Assessment of beneficial effects

(a) Types of studies considered

Several types of epidemiological study designs contribute to the evaluation of cancer prevention in humans (Table 3). These studies include experimental studies and different types of observational studies (i.e. cohort, case–control, and ecological). In addition to these types of studies, innovations in epidemiology enable other designs that may be considered in *Handbooks* evaluations.

Table 3 Types of epidemiological studies that contribute to the evaluation of cancer prevention

Experimental studies	
	• High level of investigator control over assignment to the intervention and non-intervention group • Ideally random assignment, either of individuals or of groups, to the intervention and non-intervention group • Provides evidence for the efficacy or effectiveness of a preventive intervention • Includes a range of quasi-experimental designs in which there is lack of random assignment to the intervention and non-intervention; quasi-experimental studies are often at high risk of bias
Observational (non-experimental) studies	
Cohort	• In a prospective cohort study, information on the intervention and non-intervention is collected from individuals who are then followed up over time to assess subsequent outcomes. Further intervention information may be collected at intervals during follow-up. • In a retrospective cohort study, information on intervention and subsequent outcomes in a defined group of individuals, which was usually recorded for purposes other than research, is accessed after the outcomes have occurred. • Nested within these studies, case–control and case–cohort studies provide efficiency and an opportunity to collect additional intervention information.
Case–control	• In a case–control study, individuals newly diagnosed with the outcome in a defined population and a sample of "control" individuals without the outcome from the same source population and time period are enrolled, and their intervention histories are compared. • Intervention information collected from cases and controls must refer to time before disease onset to reasonably infer a temporal association.
Mendelian randomization	• Mendelian randomization studies are cohort or case–control studies in which an intervention is inferred using appropriate genomic surrogate(s) (Yarmolinsky et al., 2018). • These studies are considered to be less prone to bias than other observational studies because the genomic variants from which intervention is inferred are randomly allocated at conception.
Ecological	• The association between an intervention and an outcome is examined not in individual people but in units of population defined geographically and/or temporally. Uncontrolled confounding is a major issue for ecological studies. • Results from ecological studies can support a hypothesis about an intervention–outcome association or, when taken together with results of case–control and cohort studies, support judgements on causal associations. • Results may be persuasive when population-wide implementation of an intervention leads to changes in cancer incidence or mortality: (a) in several populations, and there is no similar trend in similar populations not, or much less, subject to the intervention (e.g. Hakama, 1983); or (b) in a single population, by use of time series analysis when longitudinal data on both the intervention and the outcome are available (e.g. Bernal et al., 2017).

(b) Identification of eligible studies in humans

Relevant studies in humans are identified using principles of systematic review 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 a putative cancer-preventive intervention with the occurrence of cancer, or a specified intermediate outcome if it is a topic of the *Handbook*. Multiple publications on the same study population are identified so that the number of independent studies is accurately represented. Multiple publications may result, for example, from successive follow-ups of a single trial population or cohort, from analyses focused on different aspects of an intervention–outcome association, or from inclusion of overlapping populations. In these situations, the most recent or most informative report is usually reviewed first, with recourse to the other reports if important information (e.g. methodological detail) is not included in the most recent or most informative report.

(c) Study quality and informativeness

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, and experimental studies are arguably more prone, because of their smaller sample sizes, associated with the greater cost of conducting such studies.

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 intervention and an outcome. Biases that require consideration are varied and can be broadly categorized as selection bias, information bias, 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 intervention status 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 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 intervention with the outcome. 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 still arise as a result of lack of concealment, non-random allocation, lack of blinding, post-randomization exclusions, non-acceptance of or non-adherence by the study participants to the intervention condition of the study arm to which they are randomized, or study loss to follow-up. One potential shortcoming of randomized studies is their potentially limited external validity (relevance) and consequently limited generalizability to non-studied populations.

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

- **Study description:** Clarity in describing the study design, implementation, and conduct, 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 intervention and the outcome. 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 intervention of interest, and the intervention must have been assessed in a way that was not related to outcome status. In these respects, completeness of recruitment into the study from the population of interest (which is less of an issue for experimental efficacy studies than for effectiveness studies and observational studies) 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 1.2) for the 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.
- **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 in experimental studies (see Part B, Section 1.1. Of particular relevance is an assessment of the error associated with the measurement of change over time in several study designs, including prospective longitudinal studies (e.g. change in body weight estimated from contemporary recall of past body weight and self-reported or measured current body weight at recruitment into a cohort study).
- **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, as described in Part B, Section 1d) that could influence the occurrence of the outcome and may be related to the intervention of interest. 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. For example, repeated assessments of exposure to the intervention over time can be influenced by the occurrence of the outcome and thus bias the result and sometimes lead to "reverse causation". All possible sources of bias are considered for their possible impact on the results, including the possibility of reporting bias (selective reporting of some results).
- **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. For example, the use of directed acyclic graphs can inform about whether confounding and selection biases have been specified correctly (Hernán et al., 2004).

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, between the intervention and the outcome in a relevant population, 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 at least moderate heterogeneity of exposure to the intervention (intensity, frequency, and/or duration) in the study population, and biologically relevant definitions of the intervention.

(d) 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, they should comply with basic quality standards for meta-analyses and their underlying systematic reviews (e.g. AMSTAR, 2017): their risk of bias is carefully evaluated, including the completeness of the studies included, the methods used to identify and the criteria used

to select eligible studies, and the accuracy of the data extracted from 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 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.

(e) Considerations in assessing the body of epidemiological evidence

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

After the quality of individual epidemiological studies of cancer or of an intermediate outcome has been assessed and the informativeness of the various studies on the association between the intervention 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 intervention under review prevents cancer in humans. 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 false strong association. 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 outcome is common or if the intervention is highly feasible. Estimates of effects of small magnitude can also contribute useful information if the magnitude of the effect correlates with the level of 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 preventive efficacy or effectiveness than are isolated observations from single studies. If there are inconsistent results among investigations, possible reasons for such inconsistencies are sought – such as differences in time since initiation of the intervention (latency), intervention levels (e.g. dosage), or assessment methods – 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. The likelihood of reverse causation (i.e. the outcome prompts the intervention) is greater in observational studies of interventions, which often entail self-reported behaviour change, than in studies of static exposures.

An observation that cancer incidence decreases with increasing exposure to a putative preventive intervention is considered to be an indication of a preventive effect, although the absence of a graded response is not necessarily evidence against a causal relationship, and there are several reasons why the shape of the intervention–outcome association may be non-monotonic (e.g. Stayner et al., 2003).

Confidence in a causal interpretation of the evidence from studies in humans is enhanced if it is coherent with physiological and biological knowledge, including information about target organ exposure to the intervention, characteristics of tumour subtypes, and evidence of biological mechanisms by which the intervention could exert a cancer-preventive effect (see Part B, Section 4b).

The Working Group considers whether or not there are subpopulations with increased susceptibility to the cancer-preventive effects of the intervention. For example, studies that identify inter-individual differences in cancer susceptibility to the intervention on the basis of sociodemographic characteristics (e.g. age, sex, race, ethnicity), other behavioural factors (e.g. smoking or alcohol consumption), genetic polymorphisms, or age at first intervention (e.g. childhood interventions) may contribute to the identification of cancer-preventive interventions in humans. Such studies may be particularly informative if genetic polymorphisms are found to be modifiers of the intervention–outcome relationship, because evaluation of polymorphisms may increase the ability to detect an effect in susceptible subpopulations. Identifying susceptible subpopulations can also improve the specificity of targeting interventions.

2.2 Harms of the intervention

Potential harms to individuals that are linked to the intervention under review are also reviewed. Evidence of harm may come from any type of epidemiological study and may also be reported separately from evidence on the potential beneficial effects of the intervention. Although the *IARC Handbooks* do not formally evaluate the harms associated with an intervention 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.

There are three broad categories of possible harms associated with interventions: (i) biological harm (e.g. toxicity of a chemopreventive agent), (ii) physical harm (e.g. injury associated with increased physical activity), and (iii) psychosocial harm (e.g. community-based interventions and social marketing campaigns specifically targeting obesity; Walls et al., 2011). Evidence of occurrence of biological, physical, and psychosocial harm (including emerging harms identified using qualitative methods in intervention studies) is reviewed and described, and the potential impacts of the harm are discussed.

Known financial harms or opportunity costs (Walls et al., 2011), which can apply at the individual level (e.g. higher cost of healthy foods, impacts of increases in tobacco taxes on smokers of lower socioeconomic status, membership of a weight-loss plan) or the community level (e.g. community-based interventions and campaigns), may be noted.

2.3 Balance of benefits and harms

Ideally, the benefits and harms of primary prevention interventions are expressed in similar terms, such as quality-adjusted life years (QALYs) gained (benefits) or lost (harms) per 1000 individuals of the target population. After identification of all published estimates of the balance of benefits and harms based on the same combination or combinations of intervention and outcome, the Working Group selects those based on the highest-quality evaluative studies of the intervention, critically assesses each, and summarizes the results, in narrative or tabular format as appropriate. The results do not contribute to the overall evaluation of each intervention, but they may be highlighted in the rationale after the evaluation and can be used to aid decisions about implementation of and participation in the relevant primary preventive interventions.

2.4 Cost–effectiveness

For a primary preventive intervention that can deliver a beneficial outcome, cost–effectiveness is usually expressed as the estimated financial cost of implementing the intervention per unit of benefit it delivers, which is most often measured in terms of QALYs gained. 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). Although most primary preventive interventions come at a net cost to health services, some can deliver a gain in QALYs and a reduction in health service cost (Vos et al., 2010). 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.

Taking a similar approach to that taken for the balance of benefits and harms described above, the Working Group identifies published reports of well-conducted cost–effectiveness analyses based on the highest-quality evaluative studies of the primary preventive intervention, critically assesses each, and summarizes the results, in narrative or tabular format as appropriate. The results do not contribute to the overall evaluation of each intervention, but they may be highlighted in the rationale after the evaluation and can be used by governments and health services to aid decisions about implementation of the intervention for which there is sufficient evidence of a preventive effect. In addition, it is important to note that when the intervention is targeted towards a risk factor for cancer that is also a risk factor for other chronic diseases, any estimate of cost–effectiveness that is based solely on cancer is of limited use for policy purposes.

3. Studies of cancer prevention in experimental animals

(a) Types of study considered

Animal models are an important component of research on cancer prevention. Models are available that enable the evaluation of the effects of interventions on the development or progression of cancer in most major organ sites. Animal models for cancer include: (i) carcinogen-induced (e.g. chemical, physical, or infectious/biological); (ii) genetically engineered; (iii) transplantable systems (e.g. xenograft, organoid); and (iv) spontaneously developing tumours. Most cancer-preventive interventions investigated can be categorized at the biological level as those that: (i) prevent molecules from reaching or reacting with critical target sites; (ii) reduce the sensitivity of target tissues to carcinogens; or (iii) interrupt the evolution of the neoplastic process. There is increasing interest in the use of combinations of interventions as a means

of increasing efficacy and minimizing toxicity; animal models are useful in evaluating such combinations. The development of optimal strategies for intervention in humans can be facilitated by the use of animal models that mimic the neoplastic process in humans. The questions posed below (modified from Lewis et al., 2017) may assist in determining the relevance of individual studies in experimental animals to the evaluation of cancer-preventive effects in humans:

- Are the timing, route, level, and frequency of exposure comparable with those in humans, after accounting for relevant species differences?
- Is the cancer that is induced (i.e. by a biological, physical, or chemical agent, or genetic manipulation) relevant to the cancer in humans?
- Is the time at which the outcome is assessed relevant and justified?
- Does the study explore only mechanisms or pathways of cancer development?
- Is the outcome measure cancer incidence or progression rather than surrogate measures of tumour activity, such as tumour size or number of tumours?
- Do the outcome measures mimic those being evaluated in humans? More specifically, does the tumour mimic the human disease in terms of the organs or tissues affected, and at the histopathological or genetic level? Does the progression of the disease mimic the cancer in humans?

Relevant studies of cancer in experimental animals are identified using principles of systematic review as described in Part A and further detailed in the Instructions for Authors provided to each Working Group. Consideration is given to all available long-term (i.e. lifetime or near-lifetime) studies of cancer in experimental animals with the intervention under review and, when appropriate, related interventions (see Part A, Section 7). After a thorough evaluation of the pertinent study features (see Part B, Section 3b), studies judged to be irrelevant or inadequate according to the criteria determined in consultation with the Working Group may be excluded. Guidelines for conducting and reporting studies in experimental animals have been published (e.g. OECD, 2018; Percie du Sert et al., 2018).

(b) Study evaluation

Important considerations for assessing study quality include: (i) whether the intervention under review was clearly characterized; (ii) whether the intervention exposure or dose was characterized and monitored adequately; (iii) whether the control animals, exposure doses, duration of dosing, timing and frequency of dosing, duration of observation, and route of exposure to the intervention were appropriate; (iv) whether appropriate experimental animal species and strains were evaluated, including appropriate sex and age; (v) whether there were adequate numbers of animals per group; (vi) whether animals were allocated randomly to groups; (vii) whether all experimental conditions, with the exception of the tested intervention, were identical between the groups; (viii) whether the histopathology review was adequate; and (ix) whether the data were analysed correctly and reported according to well-accepted standards (e.g. Percie du Sert et al., 2018).

Specific factors to be considered in interpreting the results of cancer prevention experiments include: (i) the timing of the intervention over the course of the animals' lifespan; (ii) the timing and duration of administration of the intervention in relation to any carcinogen administration; (iii) dose–response effects; (iv) the site specificity of the anticipated cancer-preventive outcome; (v) the spectrum and relevance of the preventive outcome, from pre-neoplastic lesions to invasive cancers; (vi) the incidence, latency, and magnitude of the outcome, and the multiplicity

of the relevant neoplasms and/or other lesions; and (vii) the number and structural diversity of experimental or environmental exposures, and carcinogenic mechanisms underpinning the animals' baseline risk of the cancer to which the intervention was targeted. In addition, because administration of an intervention may result in prevention of tumours at one site but unintended consequences at other sites, it is important that multiple organs are examined in animal experiments.

Because certain factors, including diet, food or water consumption, infection, and stress, may modulate cancer risk, consideration should be given to the potential for interaction between these factors and the intervention being studied.

(c) Statistical considerations

The statistical methods used should be clearly stated and should be the generally accepted techniques refined for this purpose (Peto et al., 1980; Gart et al., 1986; Portier & Bailer, 1989; Bieler & Williams, 1993). An appropriate unit of analysis should be used (e.g. cage or individual animal in feed studies). The statistical methods should reflect the outcomes of the study (e.g. tumour incidence or multiplicity, or overall survival of the animals). For outcomes other than survival, the potential influence of different overall survival time between exposed and unexposed animals should be considered.

4. Mechanistic evidence and other relevant biological data

For a rational implementation of cancer-preventive measures, it is important not only to assess preventive end-points but also to understand the mechanisms by which the intervention exerts its cancer-preventive action. Mechanistic studies derived from human research and complemented by experimental models support cancer prevention research in humans by providing critical insight into the biological processes that can mediate the relationship between an intervention and a cancer outcome. Studies of mechanisms provide evidence for biological plausibility, inform causality, and can identify biomarkers relevant to the carcinogenic process. The study of mechanistic biomarkers can provide insights into human heterogeneity in response to carcinogens according to age, sex, genetic background, and other variables that are important to the application of cancer-preventive interventions in human populations. This array of possible contributions by mechanistic studies means that outcomes and end-points will vary widely depending on the types of intervention and the specific types of cancer examined in each *Handbook*.

Mechanistic studies and data are identified, screened, and evaluated for quality and human relevance using principles of systematic review, as described in Part A and further elaborated in the Instructions for Authors provided to each Working Group, and as detailed below.

(a) Types of studies considered

This section focuses primarily on studies in humans, including intervention trials and longitudinal studies with cancer-relevant biomarkers that may serve as exposure or intermediate end-points. Data from relevant experimental models may also be incorporated, especially when data from studies in humans are limited or are not practical to obtain.

(b) Evidence of cancer prevention

Possible mechanisms of action of interventions aiming at cancer prevention may include, but are not limited to: (i) altering the absorption, distribution, metabolism, and excretion of a known cancer-promoting or cancer-preventive agent; (ii) reducing endogenous DNA damage (e.g. by decreasing the oxidative stress and DNA–protein cross-links) or activating DNA repair or modulating epigenetic mechanisms;

(iii) altering host physiology, such as the endocrine environment (e.g. by modulation of exogenous ligands, including hormones) or the microbiome; (iv) affecting cell biology to reduce a cell's susceptibility to transformation, initiation, and progression of tumorigenesis (e.g. by regulating cell differentiation, proliferation, migration, invasion, and cell death through apoptosis and senescence); and (v) modifying the tumour microenvironment, including the inflammatory and immune responses. Inter-individual variations in these responses or outcomes associated with host factors such as age, sex, race/ethnicity, and genetic heterogeneity (e.g. metabolic polymorphisms) are also considered.

In the case of potentially chemopreventive agents, studies of absorption, distribution, metabolism, and excretion in humans and other mammalian species are summarized. The metabolic fate of the intervention agent is described, noting the metabolites that have been identified and their reactivity. A metabolic schema may indicate the relevant metabolic pathways and products, and whether supporting evidence is derived from studies in humans, in experimental animal systems, or in in vitro models. When available, physiologically based pharmacokinetic models and their parameter values are included.

(c) Harms of the preventive intervention

Any intervention that has putative beneficial effects must be assessed for potential harms. Toxic and other potentially harmful effects of a cancer-preventive intervention that are observed in studies in humans or studies in experimental animals and that might predict harmful effects in humans are reviewed, and the relevant evidence about them is summarized.

(d) Study quality and evidence synthesis

The Working Group summarizes the studies, with an emphasis on characterizing consistencies or differences in results within and across studies of varying experimental designs and model systems. Based on considerations of the quality of the studies (e.g. design, methods and reporting of results, as described in Part B, Section 3b) and relevance to humans, the Working Group may give greater weight to some included studies.

Evaluation of the results of studies in humans includes consideration of study quality, as discussed in Part B, Section 2. For observational and other studies of mechanisms of cancer prevention in humans, the quality of the study design, the intervention exposure assessment, and the accuracy (validity and precision) of the biomarker measurement are considered, as are other important factors, including those described for the evaluation of studies of cancer prevention in humans (Vermeulen et al., 2018). Specific guidelines to assess the quality of molecular biomarker and genetic studies are given in STROBE-ME (Gallo et al., 2011) and STREGA (Little et al., 2009), respectively.

In addition to studies in humans, mechanistic insights may be complemented by studies in experimental systems, including animal models (Le Magnen et al., 2016) and in vitro studies. Important considerations for in vitro studies include the ability of the system to recapitulate the carcinogenic process that occurs in humans and to model the exposure of the intervention as would be experienced in vivo (Lewis et al., 2017; Gordon et al., 2018).

The synthesis is focused on the evidence that is most informative for the overall evaluation. Evidence from several streams of mechanistic data, especially those from studies in humans, can strengthen mechanistic conclusions.

5. Summary of data reported

(a) Intervention characterization

The nature of the intervention and its characteristics, common use, and implementation in different settings, including geographical patterns and time trends, are summarized as

appropriate depending on the intervention under review. Intervention assessment methods used in key epidemiological studies reviewed by the Working Group, their strengths, and their limitations are also summarized.

(b) Cancer prevention in humans

Results of epidemiological studies pertinent to an evaluation of the cancer-preventive effects of the interventions and their harms in humans are summarized. The overall strengths and limitations of the epidemiological evidence are highlighted to indicate how the evaluation was reached. The target organ(s) or tissue(s) in which a decrease in cancer occurrence was observed are identified. Intervention–outcome associations and other quantitative data may be summarized when available. When the available epidemiological studies pertain to a mixed intervention (e.g. fruits and vegetables), the Working Group may seek to identify the specific agent or group of agents most likely to be responsible for any cancer-preventive effect. The evaluation is focused as narrowly as is appropriate or as the available data permit. Summaries of the evidence on the balance of benefits and harms and on cost–effectiveness are also provided.

(c) Cancer prevention in experimental animals

Results pertinent to an evaluation of a cancer-preventive effect in animals are summarized to indicate how the evaluation was reached. For each animal species and study design, it is stated whether or not changes in overall survival or tumour incidence, latency, severity, or multiplicity were observed, and the tumour sites are indicated. Dose–response patterns are also summarized. Possible harms of the intervention are noted.

(d) Mechanistic and other relevant data

Results pertinent to mechanisms of cancer prevention are summarized. The summary encompasses the informative studies on cancer-preventive mechanisms with adequate evidence for evaluation, and on any other aspects of sufficient importance to affect the overall evaluation. High-quality studies in humans, when available, are prioritized. In addition, supporting findings from experimental animal models or in vitro systems are summarized, especially when data from studies in humans are limited.

6. Evaluation and rationale

Evaluation of the evidence is guided by an analytical framework that depicts the relationships among the population, intervention, comparator, and outcomes (including both benefits and harms), and key contextual issues related to adherence to and implementation of the intervention and its impact on population health. The analytical framework may articulate both direct pathways (the intervention has a direct effect on cancer outcomes) and indirect pathways (the intervention has an effect on an intermediate outcome that has an established causal or preventive association with cancer incidence).

Consensus evaluations of the strength of the evidence of cancer-preventive effects of the intervention in humans, in experimental animals, and in mechanistic studies are made using transparent criteria and defined descriptive terms (see below). The Working Group then develops a consensus overall evaluation of the strength of the evidence that the intervention under review prevents cancer and assigns the intervention to one of four categories (see below).

When the Working Group has reviewed multiple, closely related interventions (e.g. different forms of an intervention on the same presumed cause of cancer), they may be grouped together for the purpose of a unified evaluation

of the strength of the evidence that they prevent cancer.

The framework for these evaluations, described below, may not encompass all factors relevant to a particular evaluation of preventive effect. After considering all relevant scientific findings, the Working Group may, exceptionally, assign the intervention to a different category from the one that a strict application of the framework would indicate, while providing a clear rationale for the overall evaluation reached.

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 prevent cancer in humans. Consideration may be given to how strongly or weakly the intervention can prevent cancer. In addition, actual and potential harms of the proposed intervention are addressed qualitatively and quantitatively, as the evidence base permits.

In what follows, the term "cancer prevention" refers to the **outcome of a *Handbooks* evaluation**, that is, to a **cancer outcome** or an **intermediate outcome**, as defined in the analytical framework. Thus, the wording of these evaluations is the same when an intermediate outcome, not cancer itself, is the outcome studied. As noted above, evaluation of an intermediate outcome is performed only when the intermediate outcome has an established causal or preventive association with cancer incidence.

(a) Cancer prevention in humans

Cancer-preventive effects in humans are evaluated on the basis of the principles outlined in Part B, Section 2. The evidence relevant to cancer prevention in humans is classified into one of the following categories:

Sufficient evidence of cancer prevention in humans: A causal preventive association between the intervention and cancer in humans has been established. That is, a cancer-preventive association has been observed consistently in the body of evidence (including several high-quality studies) and chance, bias, and confounding as causes of this association were ruled out with reasonable confidence.

Limited evidence of cancer prevention in humans: A causal preventive association between the intervention and cancer in humans is plausible. That is, a cancer-preventive association has been observed in the body of evidence, but chance, bias, or confounding as causes of this association could not be ruled out with reasonable confidence.

Inadequate evidence of cancer prevention in humans: The current body of evidence does not enable a conclusion to be drawn about the presence or absence of a preventive association between the intervention and cancer in humans. Common situations that lead to a determination of *inadequate evidence of cancer prevention in humans* include: (a) no data are available in humans; (b) there are studies available in humans, but of poor quality or informativeness; and (c) there are studies available in humans of sufficient quality, but their results are inconsistent or otherwise do not enable a conclusion to be drawn.

Evidence suggesting lack of cancer prevention in humans: There are several high-quality studies covering, through direct or indirect pathways, the full range of levels of the intervention that humans are known to encounter that are mutually consistent in not showing a preventive association between the intervention and the studied cancers at any observed level of intervention. The results from these studies alone or in combination had narrow confidence intervals with their upper bounds above or close to the

null value (e.g. a relative risk of 1.0). Similarly, bias and confounding as possible causes of this null result were ruled out with reasonable confidence, and the studies were considered informative. A conclusion of *evidence suggesting lack of cancer prevention in humans* is limited to the cancer sites, populations, life stages, conditions and levels of intervention, and length of observation covered by the pertinent studies. The target organ(s) or tissue(s) where evidence suggesting of lack of cancer prevention was observed in humans are identified.

(b) Cancer prevention in experimental animals

Cancer-preventive effects in experimental animals are evaluated on the basis of the principles outlined in Part B, Section 3. The evidence relevant to cancer prevention in experimental animals is classified into one of the following categories:

Sufficient evidence of cancer prevention in experimental animals: A preventive association has been established between the intervention and increased cancer-related survival, decreased incidence, increased latency, and/or decreased multiplicity of malignant neoplasms or of an appropriate combination of benign and malignant neoplasms in several independent, high-quality studies and model systems.

Limited evidence of cancer prevention in experimental animals: The data suggest a preventive association between the intervention and cancer in experimental animals. That is, an association has been observed but the data are limited for making a definitive evaluation because: (a) the evidence of a cancer-preventive association is based on only a few high-quality studies; (b) the intervention decreases incidence, increases latency, and/or decreases multiplicity only of benign neoplasms; or (c) there are unresolved questions about the adequacy of the design, conduct, or interpretation of the available studies.

Inadequate evidence of cancer prevention in experimental animals: The studies cannot be interpreted as showing the presence or absence of a preventive association between the intervention and cancer in experimental animals because of major qualitative or quantitative limitations of the data available, or no data are available on cancer in experimental animals.

Evidence suggesting lack of cancer prevention in experimental animals: Evidence from high-quality studies in several experimental models shows that, within the limits of the tests used (e.g. tumour site, age at intervention, conditions and levels of intervention tested), the intervention has no preventive association with cancer in experimental animals.

(c) Mechanistic evidence

Mechanistic studies are evaluated on the basis of the principles outlined in Part B, Section 4. The mechanistic evidence is classified into one of the following categories:

Strong mechanistic evidence: There are a substantial number of high-quality studies in humans that consistently link the intervention to a mechanistic pathway by which it could prevent cancer.

Limited mechanistic evidence: The evidence from mechanistic data in humans is suggestive of a cancer-preventive effect of the intervention, but (a) there are a limited number of high-quality studies, or (b) the studies cover a narrow range of experiments or relevant end-points, or (c) there are some inconsistencies in studies of similar design, or (d) there is unexplained incoherence across studies of different end-points, or (e) the available data are limited to studies in experimental model systems.

Inadequate mechanistic evidence: The evidence from mechanistic data in both humans and experimental model systems is lacking, or the data are inconsistent in linking the intervention to any mechanistic pathway by which it could prevent cancer.

(d) Overall evaluation

Finally, the body of evidence is considered as a whole. Overall evaluation of the intervention is a matter of scientific judgement that reflects the strength of the evidence derived from the studies reviewed. The levels of evidence from studies in humans, mechanistic data, and studies in experimental animals are weighed into the overall evaluation, and statements are made about cancer prevention in humans with the wording of one of the standard categories as described below.

One of the two overall evaluation scenarios (see Part A, Section 3.1) will apply, depending on the nature of the evidence that has been reviewed (Table 4; see also Part A). If, for logistic reasons, evidence for Step 1 and Step 2 of Scenario 2 has been reviewed at two separate *Handbooks* meetings, no overall evaluation will be made for Step 2 alone.

None of these evaluations quantify the fraction of the burden of a particular cancer that a specific intervention would prevent; thus, some interventions may prevent a small fraction of the cancer, some may prevent a larger fraction, and these fractions may vary across populations, for example as a function of the prevalence of the relevant risk factors.

Overall evaluation categories

(i) The intervention *is established to prevent cancer in humans* (Group A)

This category is used for interventions for which there is *sufficient evidence* of cancer prevention in humans, either directly (Scenario 1) or in two steps (Scenario 2): from the intervention to the intermediate outcome (Step 1) and from the intermediate outcome to cancer (Step 2).

The organ sites on which the evidence in humans is based are stated here. A statement is also made of what the Working Group considers to be the magnitudes of the benefits and the harms of the intervention, in as nearly comparable terms as possible, for people adhering to the intervention as commonly implemented in practice, and whether or not the benefits outweigh the harms.

(ii) The intervention *probably prevents cancer in humans* (Group B1)

In Scenario 1, this category is used for interventions for which there is *limited evidence* of cancer prevention in humans and either *strong mechanistic evidence* in humans or *sufficient evidence* in experimental animals with all the criteria for the relevance to humans being met (see Part B, Section 3a).

In Scenario 2, this category is used for interventions for which there is *sufficient evidence* in humans that the intervention has a cancer-preventive effect on the intermediate outcome (Step 1), *limited evidence* that the intermediate outcome has a cancer-preventive effect in humans (Step 2), and either *sufficient evidence* in experimental animals with all the criteria for the relevance to humans being met or *strong mechanistic evidence* in humans (see Part B, Section 3a). Alternatively, this category is used when there is *limited evidence* in humans that the intervention has a cancer-preventive effect in the intermediate outcome (Step 1) and *sufficient evidence* that the intermediate outcome has a cancer-preventive effect in humans (Step 2).

(iii) The intervention *possibly prevents cancer in humans* (Group B2)

In Scenario 1, this category is used for interventions for which there is *limited evidence* of cancer prevention in humans, *less than strong evidence* from mechanistic data, and *less than sufficient evidence* of cancer prevention in experimental animals.

In Scenario 2, this category is used when (i) there is *sufficient evidence* in humans that the intervention has a cancer-preventive effect on the intermediate outcome (Step 1), and *limited evidence* in humans and *less than sufficient evidence* in experimental animals or *less than strong evidence* from mechanistic data that the intermediate outcome has a cancer-preventive

Table 4 Summary of the strength of the evidence in each evidence stream contributing to the overall evaluation

Scenario 1: Direct evidence that the intervention prevents cancer			
Strength of the evidence that the intervention prevents cancer in humans	**Strength of the evidence from mechanistic studies that the intervention prevents cancer**	**Strength of the evidence that the intervention prevents cancer in experimental animals**	**Overall evaluation**
Sufficient	–	–	Group A
Limited	*Strong*	–	Group B1
Limited	–	*Sufficient*	Group B1
Limited	*Less than strong*	*Less than sufficient*	Group B2
Inadequate	–	–	Group C
Evidence suggesting lack of cancer prevention	–	*Evidence suggesting lack of cancer prevention*	Group D

Scenario 2: Evidence that the intervention prevents cancer by way of an intermediate outcome (risk factor or preventive factor)			
Step 1	**Step 2**[a]		**Overall evaluation**[a]
Strength of the evidence that the intervention decreases exposure to the risk factor or increases exposure to the preventive factor in humans	**Strength of the evidence that decreasing exposure to the risk factor or increasing exposure to the preventive factor prevents cancer in humans**	**Strength of the evidence that decreasing exposure to the risk factor or increasing exposure to the preventive factor prevents cancer in experimental animals or mechanistic studies**[b]	
Sufficient	*Sufficient*[c]	–	Group A
Sufficient	*Limited*	*Sufficient*	Group B1
Sufficient	*Limited*	*Less than sufficient*	Group B2
Limited	*Sufficient*	–	Group B1
Limited	*Limited*	–	Group B2
Inadequate	–	–	Group C
–	*Evidence suggesting lack of cancer prevention*	*Evidence suggesting lack of cancer prevention*	Group D
Evidence suggesting lack of cancer prevention	–	–	Group D

[a] This overall evaluation applies only when evidence from both Step 1 and Step 2 is available. When a *Handbook* evaluates only Step 2, no overall evaluation is made.

[b] Evidence in experimental animals and mechanistic data is considered to be *sufficient* when there is *strong evidence* from mechanistic data (mechanistic studies in humans) or *sufficient evidence* in experimental animals.

[c] The evidence in this category may be considered to be *sufficient* when it is based on observational studies of change in cancer incidence associated with self-reported or observed (by way of time-separated repeated measures) change in the level of a risk factor or preventive factor (e.g. smoking cessation; increase in consumption of fruits and vegetables), OR, exceptionally, studies of variation in cancer incidence with the level of a risk factor or preventive factor measured at one time point.

effect; OR (ii) there is *limited evidence* in humans that the intervention has a cancer-preventive effect on the intermediate outcome (Step 1), and *limited evidence* in humans that the intermediate outcome has a cancer-preventive effect, and any evidence category in experimental animals and mechanistic data.

When the evidence is classified in Group B1 or Group B2, the evaluation is followed by a description of harms, actual and potential.

(iv) The intervention *is not classifiable as to its capacity to prevent cancer in humans* (Group C)

In both Scenario 1 and Scenario 2, this category is used for interventions for which there is *inadequate evidence* in humans, irrespective of the level of evidence from mechanistic data and studies in experimental animals. Interventions that do not fall into any other category are also placed in this category.

(v) The intervention *probably does not prevent cancer in humans* (Group D)

In Scenario 1, this category is used for interventions for which there is *evidence suggesting lack of cancer prevention* both in humans and in experimental animals. In Scenario 2, this category is used when there is *evidence suggesting lack of cancer prevention* both in humans and in experimental animals for the intermediate outcome to cancer, irrespective of the level of evidence for the intervention to the intermediate outcome; or there is *evidence suggesting lack of cancer prevention* for the intervention to the intermediate outcome, irrespective of the level of evidence for the intermediate outcome to cancer.

(e) Rationale

The reasoning that the Working Group used to reach its evaluation is summarized so that the basis for the evaluation offered is transparent. It includes concise statements of the principal line or lines of argument that emerged in the deliberations of the Working Group, the conclusions of the Working Group on the strength of the evidence for each stream, 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 evaluations.

In the rationale, the Working Group may draw attention to the fact that actions on the evaluations should be taken in the light of country- or setting-specific circumstances that influence the public health priority, feasibility, and acceptability of programmes based on the interventions evaluated.

References

Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, et al. (2016). Mutational signatures associated with tobacco smoking in human cancer. *Science*. 354(6312):618–22. doi:10.1126/science.aag0299 PMID:27811275

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. doi:10.1002/ajim.4700130107 PMID:3344750

Bernal JL, Cummins S, Gasparrini A (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 46(1):348–55. doi:10.1093/ije/dyw098 PMID:27283160

Bieler GS, Williams RL (1993). Ratio estimates, the delta method, and quantal response tests for increased carcinogenicity. *Biometrics*. 49(3):793–801. doi:10.2307/2532200 PMID:8241374

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

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. doi: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. doi:10.1289/ehp.6950 PMID:15345338

Gallo V, Egger M, McCormack V, Farmer PB, Ioannidis JP, Kirsch-Volders M, et al. (2011). STrengthening the Reporting of OBservational studies in Epidemiology - Molecular Epidemiology (STROBE-ME): an extension of the STROBE statement. *Prev Med*. 53(6):377–87. doi:10.1016/j.ypmed.2011.08.007 PMID:22029945

Gart JJ, Krewski D, Lee PN, Tarone RE, Wahrendorf J (1986). Statistical Methods in Cancer Research, Vol. 3. The design and analysis of long-term animal experiments. Lyon, France: International Agency for Research on Cancer (IARC Scientific Publications, No. 79)

Gordon JL, Brown MA, Reynolds MM (2018). Cell-based methods for determination of efficacy for candidate therapeutics in the clinical management of cancer. *Diseases*. 6(4):E85. doi:10.3390/diseases6040085 PMID:30249005

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.

Hakama M (1983). Trends in the incidence of cervical cancer in the Nordic countries. In: Magnus K, editor. Trends in cancer incidence: causes and practical implications. Taylor & Francis Inc.; pp. 279–92.

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

Hill AB (1965). The environment and disease: association or causation? *Proc R Soc Med*. 58(5):295–300. doi:10.1177/003591576505800503 PMID:14283879

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

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

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

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

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

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

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

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

IARC (2018). Guidelines for Observers at *IARC Handbooks* Meetings. Available from: https://handbooks.iarc.fr/preambles/guidelines-for-observers/.

IARC (2019). 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.

Jonas DE, Ferrari RM, Wines RC, Vuong KT, Cotter A, Harris RP (2018). Evaluating evidence on intermediate outcomes: considerations for groups making health-care recommendations. *Am J Prev Med*. 54(1S1):S38–52. doi:10.1016/j.amepre.2017.08.033 PMID:29254524

Le Magnen C, Dutta A, Abate-Shen C (2016). Optimizing mouse models for precision cancer prevention. *Nat Rev Cancer*. 16(3):187–96. doi:10.1038/nrc.2016.1 PMID:26893066

Lewis SJ, Gardner M, Higgins J, Holly JMP, Gaunt TR, Perks CM, et al. (2017). Developing the WCRF International/University of Bristol methodology for identifying and carrying out systematic reviews of mechanisms of exposure-cancer associations. *Cancer Epidemiol Biomarkers Prev*. 26(11):1667–75. doi:10.1158/1055-9965.EPI-17-0232 PMID:28978562

Little J, Higgins JP, Ioannidis JP, Moher D, Gagnon F, von Elm E, et al. (2009). STrengthening the REporting of Genetic Association studies (STREGA) – an extension of the STROBE statement. *Eur J Clin Invest*. 39(4):247–66. doi:10.1111/j.1365-2362.2009.02125.x PMID:19297801

OECD (2018). Test no. 451: carcinogenicity studies. OECD Guidelines for the Testing of Chemicals, Section 4. Paris, France: OECD Publishing.

Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Altman DG, Avey MT, et al. (2018). Revision of the ARRIVE guidelines: rationale and scope. *BMJ Open Sci*. 2(1):e000002. doi:10.1136/bmjos-2018-000002 PMID:33954268

Peto R, Pike MC, Day NE, Gray RG, Lee PN, Parish S, et al. (1980). Guidelines for simple, sensitive significance tests for carcinogenic effects in long-term animal experiments. *IARC Monogr Eval Carcinog Risk Chem Hum Suppl*. (2 Suppl):311–426. PMID:6935185

Porta M, editor (2008). A dictionary of epidemiology. 5th ed. New York (NY), USA: Oxford University Press. doi:10.1093/acref/9780195314496.001.0001

Porta M, editor (2014). A dictionary of epidemiology. 6th ed. New York (NY), USA: Oxford University Press. doi:10.1093/acref/9780199976720.001.0001

Portier CJ, Bailer AJ (1989). Testing for increased carcinogenicity using a survival-adjusted quantal response test. *Fundam Appl Toxicol*. 12(4):731–7. doi:10.1016/0272-0590(89)90004-3 PMID:2744275

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. doi:10.1093/aje/kwu211 PMID:25245043

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

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.

Stayner L, Steenland K, Dosemeci M, Hertz-Picciotto I (2003). Attenuation of exposure-response curves in occupational cohort studies at high exposure levels. *Scand J Work Environ Health*. 29(4):317–24. doi:10.5271/sjweh.737 PMID:12934726

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

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

Thun MJ, Linet MS, Cerhan JR, Haiman CA, Schottenfeld D, editors (2018). Cancer epidemiology and prevention. 4th ed. New York (NY), USA: Oxford University Press.

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. doi:10.1093/ije/dyv341 PMID:26800751

Vermeulen R, Bell DA, Jones DP, Garcia-Closas M, Spira A, Wang TW (2018). Application of biomarkers in cancer epidemiology. In: Thun MJ, Linet MS, Cerhan JR, Haiman CA, Schottenfeld D, editors. Cancer epidemiology and prevention. 4th ed. New York (NY), USA: Oxford University Press; pp. 77–96.

Vos T, Carter R, Barendregt J, Mihalopoulos C, Veerman JL, Magnus A, et al.; ACE–Prevention Team (2010). Assessing Cost-Effectiveness in Prevention (ACE–Prevention): final report. Brisbane, Australia: University of Queensland and Melbourne, Australia: Deakin University. Available from: https://public-health.uq.edu.au/files/571/ACE-Prevention_final_report.pdf.

Walls HL, Peeters A, Proietto J, McNeil JJ (2011). Public health campaigns and obesity - a critique. *BMC Public Health*. 11(1):136. doi:10.1186/1471-2458-11-136 PMID:21352562

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

Yarmolinsky J, Wade KH, Richmond RC, Langdon RJ, Bull CJ, Tilling KM, et al. (2018). Causal inference in cancer epidemiology: what is the role of Mendelian randomization? *Cancer Epidemiol Biomarkers Prev*. 27(9):995–1010. doi:10.1158/1055-9965.EPI-17-1177 PMID:29941659