

Working procedures

Starting in 2006, the series of International Agency for Research on Cancer (IARC) Handbooks of Cancer Prevention added tobacco control as a new area of prevention for their reviews. When appropriate, in addition to cancer, other health outcomes preventable by avoiding tobacco use or exposure to secondhand smoke (SHS) may be included when evaluating the effectiveness of interventions in a Handbook.

The Working Procedures described herein are largely taken from the Handbooks of Cancer Prevention devoted to Chemoprevention and Screening, and from the IARC Monograph Preamble (updated in January 2006).

The text that follows is organised in two principal parts. The first addresses the general scope, objectives, and structure of the Handbooks with emphasis on tobacco control. The second describes the scientific procedures for evaluating cancer-preventing agents and tobacco control policy interventions.

The term “exposure” appears repeatedly in these procedures, borrowed from the IARC Monographs devoted to the evaluation of carcinogenicity. Epidemiological studies conducted to assess the association between exposure to a

given hazard and disease outcome, are based on the meaning of the term “exposure” implying increased risk to an undesired health effect. However, in this series of Handbooks, dedicated to the evaluation of the preventive effects of compounds, biological or pharmaceutical products, behaviours, programmes, and policy interventions, the traditional meaning of the term “exposure” is unfitting. Therefore in several instances the term “intervention,” which lacks a hazardous connotation, is preferred. Examples of interventions with expected benefits in the area of tobacco control are tobacco use cessation, banning of smoking in public places, and taxation on tobacco products.

Part one: General principles

General scope

The prevention and control of cancer are the strategic objectives of IARC. Cancer prevention may be achieved at the individual level by avoiding cancer-causing agents (e.g. not using tobacco products), and at the population level by adopting programmes or legislation to reduce or eliminate exposure to cancer-causing agents (e.g. removing exposure to SHS through banning smoking in public and workplaces).

The Handbooks on tobacco control will evaluate the strength of the available evidence on the effects of interventions intended to prevent or reduce tobacco use, tobacco supply, and, when possible, tobacco-associated morbidity and mortality. The aim of the Handbook series is to provide the scientific community, policymakers, and governing bodies of IARC member states, as well as other countries with evidence-based assessments of these interventions at the individual and population levels, with the ultimate goal of assisting in the global implementation of tobacco control provisions within national and international programmes aimed at reducing tobacco-related morbidity and mortality.

Objectives

The objective is to prepare and publish, in the form of Handbooks, critical reviews and consensus evaluations of evidence on the effects of interventions focusing on tobacco control, with the help of an internationally formed Working Group (WG) of experts. The Handbooks may also indicate where additional research efforts are needed, specifically when data immediately relevant to an evaluation are not available. The evaluations in the Handbooks are scientific

and qualitative judgments of peer-reviewed, published data, conducted during a six-day meeting of peer review and discussions by the WG.

Topic for the Handbook

The topic to be evaluated in a Handbook is selected approximately 12 months prior to the meeting by the head of the Lifestyle, Environment, and Cancer Group, after consultation with IARC scientists involved in tobacco research. A Handbook may cover a single topic or a group of related topics in the area of tobacco control.

Meeting participants

Soon after the topic of a Handbook is chosen, international scientists with relevant expertise are identified by IARC staff (usually through literature searches), in consultation with other experts. Each participant serves as an independent scientist and not as a representative of any organisation, government, or industry. Every effort is made to achieve a balanced group of experts in terms of gender, geographic origin, expertise, and diversity of scientific opinion.

Five categories of participants may attend Handbook meetings: WG members, Invited Specialists, Representatives of national and international health agencies, Observers, and the IARC Secretariat. Participants in the first two groups generally have published significant research related to the topic being reviewed or in tobacco control in particular. All participants are listed, with their addresses and principal affiliations, at the beginning of each Handbook volume. A description of each participant type, and their responsibilities, is listed below.

1. *The Working Group* is responsible for the critical

reviews and evaluations that are developed during the meeting. WG members are selected based on knowledge and experience pertinent to the topic evaluated and absence of real or apparent conflicts of interest. The tasks of the WG are: (i) to ascertain that all appropriate data have been collected; (ii) to select the data relevant for the evaluation on the basis of scientific merit; (iii) to prepare accurate summaries of the data to enable the reader to follow the reasoning of the WG; (iv) to critically evaluate the results of epidemiological, clinical, and other type of studies; (v) to prepare recommendations for research and for public health action; and (vi) if the topic being reviewed so permits, to make an overall evaluation of the evidence of a protective effect or reduced risk associated with the exposure or intervention focus of the evaluation.

2. *Invited Specialists* are experts who also have critical knowledge and experience, but have a real or apparent conflict of interest. These experts are invited, when necessary, to assist in the WG by contributing their unique knowledge and experience during subgroup and plenary discussions. They may also contribute text on the intervention being evaluated. Invited Specialists do not serve as meeting chair or subgroup chair, redact summaries, or participate in the evaluations.

3. *Representatives* of national and international health agencies may attend meetings because their agencies are interested in the topic of a Handbook. Representatives do not serve as meeting chair or subgroup chair, draft any part of a Handbook, or

participate in the evaluations.

4. *Observers* with relevant scientific credentials may be admitted to a meeting by IARC in limited numbers. Priority will be given to achieving a balance of Observers from constituencies with differing perspectives. They are invited to observe the meeting and should not attempt to influence it. Observers serve as sources of first-hand information from the meeting to their sponsoring organisations. They can play a valuable role in ensuring that all published information and scientific perspectives are considered. Observers will not serve as meeting chair or subgroup chair, draft any part of a Handbook, or participate in the evaluations. At the meeting, the chair and subgroup chairs may grant Observers the opportunity to speak, generally after they have heard a discussion.

5. The *IARC Secretariat* consists of scientists who have relevant expertise and are designated by IARC to attend a meeting. They serve as rapporteurs and participate in all discussions. When requested by the meeting chair or subgroup chair, they may also draft text or prepare tables and analyses.

The WHO Declaration of Interest form is sent to each prospective participant at the first contact, with the preliminary letter presenting the Handbook meeting. Before an official invitation is extended, each potential participant, including the IARC Secretariat, completes the WHO Declaration of Interest form to report financial interests, employment and consulting, and individual and institutional research support related to the topic of the meeting. IARC assesses the declared interests to determine whether there is a conflict

that warrants some limitation on participation. WG members are selected based on the absence of real or apparent conflicts of interest. If a real or apparent conflict of interest is identified, then the expert is asked to attend as an Invited Specialist. The declarations are updated and reviewed again at the opening of the meeting, approximately eight months later. Interests related to the subject of the meeting are disclosed to the meeting participants and in the published volume (Cogliano *et al.*, 2004).

Data for the Handbooks

The Handbooks review all pertinent studies on the intervention to be evaluated. Only those data considered relevant to evaluate the evidence are included and summarised. Those judged inadequate or irrelevant to the evaluation may be cited but not summarised. If a group of similar studies is not reviewed, the reasons are indicated.

With regard to reports of basic scientific research, epidemiological studies, and clinical trials, only studies that have been published or accepted for publication in the openly available scientific literature are reviewed. In certain instances, government agency reports that have undergone peer review and are widely available can be considered. Exceptions may be made ad hoc to include unpublished reports that are in their final form and publicly available, if their inclusion is considered pertinent to making an evaluation. Abstracts from scientific meetings, and other reports that do not provide sufficient detail upon which to base an assessment of their quality are generally not considered.

Inclusion of a study does not imply acceptance of the adequacy of the study design or of the analysis

and interpretation of the results, and limitations identified by the WG are clearly outlined in square brackets (i.e. []). The reasons for not giving further consideration to an individual study are also indicated in square brackets. Important aspects of a study, directly impinging on its interpretation, are brought to the attention of the reader. In general, numerical findings are indicated as they appear in the original report; units are converted when necessary for easier comparison. The WG may conduct additional analyses of the published data and use them in their assessment of the evidence. These analyses and their results are outlined in square brackets or in italics in the Handbook.

Working procedures

Chair of the meeting

The chair of the Handbook meeting is identified among leading international experts soon after the topic of a Handbook is chosen. The chair will help develop an outline for the Handbook early on and aid in identifying prospective experts to form the WG. The chair participates on conference calls with WG members and Invited Specialists in preparing for the meeting, provides early feedback on working papers, directs the meeting, and helps resolve queries emerging on the working papers once the meeting is over.

Literature to be reviewed

After the topic of the Handbook is chosen and invited experts have accepted to take part in the Handbook meeting, participants are expected to conduct their own literature searches in support of their working assignments. Only upon expressed

request by any of the members of the WG, pertinent studies are identified by IARC from recognised sources of information, such as PubMed, and made available to WG members and Invited Specialists to prepare the working papers for the meeting. Paper and/or electronic version of the studies cited in the working papers are available at the time of the meeting for consultation by the WG and all other participants including the IARC Secretariat.

Working papers

Working papers are due about six to eight months after original contact of invited experts. The first version of the working papers is compiled and formatted by IARC staff about two months prior to the meeting, or as soon as they are received, and made available ahead of time through IARC's internet to all WG members, Invited Specialists, and the IARC Secretariat. Reception of working papers ahead of the established deadline is encouraged, as it allows review of their content, facilitating identification of information gaps from the start. When possible, or when deemed necessary, working papers may be discussed early on among experts to expedite the review process to be accomplished during the meeting. Conference calls will be scheduled after reception of all working papers and prior to the meeting, with the aim of identifying areas deserving additional work by experts before the meeting.

Acknowledgement of significant contributions to the chapters by colleagues of the invited experts, either at their home institution or elsewhere, can be included in the Handbook under an acknowledgement paragraph to be shown following the listing of the meeting participants.

Meeting

The meeting participants convene at IARC for six days to discuss and finalise the texts of the working papers that will constitute the Handbook and to formulate the evaluations. The WG members and Invited Specialists are grouped into subgroups according to their area of expertise. Subgroups meet during the first three to four days to review in detail the first versions of their working papers, develop a joint subgroup draft, and write summaries. During the last few days the participants meet in plenary session to review the subgroup working papers and summaries and to develop the consensus evaluations. Scheduling of plenary and subgroup time may change from one Handbook meeting to another.

Post-meeting

After the meeting, the draft Handbook is verified by consulting the original literature, edited, and prepared for publication by IARC staff. The aim is to publish Handbooks within 12-15 months of the meeting. If applicable, summaries reporting the results of the evaluation may be available on the IARC website (<http://www.iarc.fr>) soon after the meeting, and a short report may be published in the international literature.

Part two: Scientific review of the evidence and evaluation

Scientific review

The evidence forming the foundation of the evaluation results from the studies reviewed. The validity of these studies will be examined critically to determine the weight they contribute to the assessment. This entails judging the appropriateness of study design, data collection

(including adequate description of the intervention and follow-up), data analysis, and ultimately, deciding if chance, bias, confounding, or lack of statistical power may account for the observed results. The experts will ascertain how the limitations of the studies affect the results and conclusions reported. The criteria that follow apply to epidemiological and clinical studies, and therefore may not be as relevant to studies where other quality criteria would be indicated (e.g. those assessing the impact of economic policies or when health outcomes are not contemplated).

Quality of studies considered

It is necessary to take into account the possible roles of bias, confounding, and chance in the interpretation of epidemiological studies. Bias is the operation of factors in the study design or execution that leads erroneously to a stronger or weaker association than in fact exists between the exposure/intervention being evaluated and the outcome. Confounding is a form of bias that occurs when the association with the disease is made to appear stronger or weaker than it truly is, as a result of an association between the apparent causal factor and another factor that is associated with either an increase or decrease in the incidence of the disease. The role of chance is related to biological variability and the influence of sample size on the precision of estimates of effect.

In evaluating the extent to which these factors have been taken into account in an individual study, the Handbook considers a number of aspects of design and analysis as described in the report of the study.

First, the study population, disease (or diseases), and exposure/intervention should have been well-

defined by the authors. Cases of disease in the study population should have been identified independently of the intervention of interest, and the intervention assessed in a way that was not related to disease status.

Second, in the study design and analysis, the authors should have taken into account other variables that can influence the risk of disease or impact of an intervention and that may have been related to the intervention of interest. Potential confounding by such variables should have been dealt with either in the design of the study, such as by matching, or in the analysis, by statistical adjustment. In cohort studies, comparisons with local rates of the disease may or may not be more appropriate than those with national rates. Internal comparisons of disease frequency among individuals at different levels of the intervention are also desirable in cohort studies, since they minimise the potential for confounding related to difference in risk factors between an external reference group and the study population.

Third, the authors should have reported the basic data on which the conclusions are founded, even if sophisticated statistical analyses were employed. The numbers of exposed and unexposed cases and controls in a case-control study, and the numbers of cases observed and expected in a cohort study should have been provided. Further tabulations by time since exposure began, and other temporal factors, are also important. In a cohort study, data on all cancer sites and all causes of death should have been given to reveal the possibility of reporting bias. In a case-control study, the effects of investigated factors other than the exposure of interest should have been reported.

Finally, the statistical methods used to obtain estimates of relative

risk, absolute rates of cancer, confidence intervals, and significance tests, and to adjust for confounding should have been clearly stated by the authors.

Aspects that are particularly important in evaluating experimental studies are: the selection of participants, the nature and adequacy of the randomisation procedure, evidence that randomisation achieved an adequate balance between groups, the exclusion criteria used before and after randomisation, compliance with the intervention in the intervention group, and 'contamination' with the intervention in the control group. Other considerations are the means by which the endpoint was determined and validated, the length and completeness of follow-up of the groups, and the adequacy of the analysis.

Detailed analyses of both relative and absolute risks in relation to temporal variables, such as age at first exposure, time since first exposure, duration of exposure, cumulative exposure, peak exposure (when appropriate), and time since exposure ceased, will be reviewed and summarised when available.

Independent, population-based studies of the same exposure or intervention may lead to ambiguous results. Combined analyses of data from multiple studies may be a means of resolving this ambiguity. There are two types of combined analysis: the first combines summary statistics, such as relative risks, from individual studies (meta-analysis); the second involves a pooled analysis of the raw data from the individual studies (pooled analysis).

Advantages of combined analyses include better precision due to increased sample size, as well as the opportunity to explore potential confounders, interactions, and modifying effects that may

explain heterogeneity among studies in more detail. A disadvantage of combined analyses is the possible lack of compatibility of data from various studies due to differences in subject recruitment, data collection procedures, measurement methods, and effects of unmeasured covariates that may differ between studies.

Meta-analyses may be conducted by the WG during the course of preparing a Handbook and are identified as original calculations by placement of the results in square brackets or in italics. These may be de novo analyses or updates of previously conducted analyses that incorporate the results from new studies. Whenever possible, however, such analyses are preferably conducted preceding the Handbook meeting. Publication of the results of such meta-analyses prior to, or concurrently with, the Handbook meeting is encouraged for purposes of peer review. The same criteria for data quality that would be applied to individual studies must be applied to combined analyses, and such analyses must take into account heterogeneity between studies.

Criteria for causality

After the quality of each study has been summarised and assessed, a judgment is made concerning the strength of evidence that the exposure or intervention in question reduces the risk of disease or is protective for humans. Hill (1965) lists areas for evaluating the strength of epidemiological associations used in the review of human data when assessing carcinogenesis. These criteria, in many instances, will apply to the assessment included in a Handbook:

- Consistency of observed associations across studies and populations;

- Magnitude of the reported association;
- Temporal relationship between exposure/intervention and change in disease;
- Exposure-response biologic gradient;
- Biological plausibility;
- Coherence of results across other lines of evidence; and
- Analogy present in related exposures and their effects on health.

If the results are inconsistent among investigations, possible explanations are sought (e.g. differences in level of exposure/intervention). Results of studies judged to be of high quality are given more weight than those of studies judged to be less methodologically sound. When several studies show little or no indication of an association between an intervention and cancer prevention, the judgment may be made that, in the aggregate, they show evidence of lack of effect. The possibility that bias, confounding, or misclassification of exposure or outcome could explain the observed results should be considered and excluded when reasonable certainty exists.

Assessing studies reporting the impact of tobacco control policy interventions not necessarily contemplating health outcomes

Evaluating the outcomes of population level tobacco control policy involves three interrelated questions: (1) Does the policy have an impact? (causality); if so, (2) Under what conditions? (moderation); and (3) How (mediation)?

The choice of design elements will depend on which questions are considered to be a part of the evaluation effort. It is important to ensure that the appropriate concepts

are chosen, and, that for each, measures are identified that are suitable to answer the evaluation question.

In the absence of a randomised trial, there are two study design strategies that can be employed for the rigorous evaluation of the effects of policies. First is the use of measurements both before and after the policy's implementation. These measurements can be taken from either units (usually, but not limited to, individuals; the same logic would apply if the measures were of households, schools, or other venues) that are either the same (as in a cohort design) or different, but drawn from the same sampling process (as in a repeat cross-sectional design). The second design strategy is the use of a quasi-experimental design, in which one group that is exposed to a policy is compared to a similar unexposed group, as discussed above. Combining these two strategies in a single study yields a two-group, pre-post design, which offers a higher degree of internal validity than either feature alone. The utility of longitudinal designs is strengthened if there are multiple data collections before and/or after policy implementation, allowing more precise specification of effects (e.g. taking into account temporal trends that were occurring before the implementation of the policy).

A distinction between study designs and study features is worth noting. In addition to the two design considerations stated above, there are two study feature strategies that contribute to increasing an evaluation study's internal validity. The first is the measurement of policy-specific variables that are theorised to be affected initially after the policy is implemented. A second strategy is the measurement of policy-specific variables for policies that

have not changed; such variables act as another form of control. Recommendations for measures pertinent to the evaluation of each WHO Framework Convention on Tobacco Control policy domain are provided in Handbook Volume 12 (IARC, 2008).

Combining the two design and two study feature strategies, along with the inclusion of other explanatory variables (covariates) that might help explain differences between two jurisdictions, creates a powerful research design, allowing more confident inferences to be made about the causal effects of policies and/or combinations of policies.

Evaluation efforts should be informed by knowledge of the nature of the policy being evaluated, and the goals of the evaluation study should be clearly stated. Evaluation planning should be guided by understanding what threats to internal validity may be present in the study of a given policy situation, and then adding design elements and other measures to reduce or eliminate those threats.

Knowledge of the mediational pathways that are theorised to explain how policy affects behaviour and environment (or environmental risk) should lead to an appropriate study design, the inclusion of appropriate constructs and measures, and the selection of analytic tools that are well-suited to estimating the causal impact of policies by providing an explanatory pathway and helping to eliminate alternative explanations.

The utility of longitudinal designs is strengthened if there are multiple data collections before and/or after policy implementation, as this allows more precise specification of effects (e.g. taking into account temporal trends that were occurring before the implementation of the policy). The role of time series analysis on aggregate sales/consumption data

that demonstrate the effect of price on consumption is a good example of the power of multiple measurements.

Both repeated cross-sectional and longitudinal (cohort) designs are useful for assessing the impact of a given policy. The use of cohort designs provides an additional capability for tracking the impact of policies within individuals, allowing stronger tests of mediational pathways.

Addition of samples from other populations to either or both intervention and control arms, also adds strength to the evaluation design, as does having varying levels of intensity of the intervention.

Similarly, parallel assessment of alternative explanations for observed changes in outcomes (e.g. possibly being due to other policies or industry counter-actions) adds strength over assessing these effects in separate studies.

The existence of studies with complementary strengths and weaknesses is particularly useful in triangulating the results of a corpus of evaluation studies to see if a consistent pattern emerges.

The use of probability sampling in an evaluation study increases its external validity - the extent to which the findings of a policy evaluation study can be generalised to draw conclusions about the impact of the policy on the larger population.

At a broader level, the design of an evaluation study should be guided by knowledge of how prior evaluation studies in the same policy domain have been conducted. An analysis of the similarity or differences in policy impact across similar studies can yield powerful conclusions about the overall impact of a policy.

Summary of the data reviewed (evidence)

This section summarises in a concise manner the results of the evidence presented in the preceding sections in a Handbook. Traditionally, this section does not include citation of literature, as do preceding sections presenting and discussing the evidence.

Evaluation of the evidence

An evaluation of the strength of the evidence for disease prevention or reduction in morbidity and mortality is made using standard terms described in previous volumes of the Handbooks of Cancer Prevention (e.g. Volume 11). In evaluating the strength of the evidence on the effects of tobacco control interventions directed at the population, disease prevention or health outcomes may not always be a measurable endpoint. Also, it is conceivable that not every exposure/intervention reviewed in a Handbook of tobacco control will permit a formal evaluation of the evidence, as traditionally done in other Handbooks of Cancer Prevention and in the Monographs.

The following criteria are proposed when evaluating the weight of the evidence on the effects of tobacco control interventions:

Sufficient evidence: The WG considers that an association has been observed in studies in which chance, bias, and confounding can be ruled out with reasonable confidence. The association is highly likely to be causal. A statement that there is *sufficient evidence* should be followed by a separate sentence that identifies the nature and magnitude of the observed effect.

Strong evidence: There is consistent evidence of an association between the

intervention under consideration and a given effect, but evidence of causality is limited by the fact that chance, bias, or confounding have not been ruled out with reasonable confidence. However, explanations other than causality are unlikely.

Limited evidence: There is some evidence of association between the intervention under consideration and a given effect, but alternative explanations are possible.

Inadequate/no evidence: There are no available methodologically sound studies showing an association. The available studies are of insufficient quality, consistency, or statistical power to permit a conclusion regarding the presence or absence of a causal association between the intervention and a given effect. Alternatively, this category is used when no data are available.

Evidence suggesting lack of effect: There are several methodologically adequate studies that are mutually consistent in not showing an association between the intervention and a given effect.

Overall evaluation

The overall evaluation, usually in the form of a narrative, will include a summary of the body of evidence considered as a whole, and summary statements made about the strength of the evidence for policy effects, including changes in tobacco use, changes in health risks, and incidental effects.

IARC WGs make every effort to achieve a consensus evaluation. Consensus reflects broad agreement among WG members, but not necessarily unanimity. The chair may elect to poll WG members to

determine the diversity of scientific opinion on issues where consensus is not readily apparent.

Recommendations

After reviewing the data and deliberating on them, the WG may formulate recommendations, where applicable, for further research and public health action.

References

Cogliano VJ, Baan RA, Straif K et al. (2004). The science and practice of carcinogen identification and evaluation. *Environ Health Perspect*, 112:1269–1274. doi:10.1289/ehp.6950 PMID:15345338

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

IARC (2008). IARC Handbook of Cancer Prevention, Tobacco Control, Vol.12: Methods for Evaluating Tobacco Control Policies.

