Table of contents

Public Comments Form

To Propose an Update to the Working Procedures of the *IARC Handbooks of Cancer Prevention*

Please send completed comment forms to hbworkingproc@iarc.fr.

1. Name and affiliation of commenter

Your name	David B Thomas
Your principal affiliation	Fred Hutchinson Cancer Research Center (retired)
If another party suggested that you submit this nomination, please identify	Click here to enter text.
WHO Declaration of Interests form (to sign and submit via hbworkingproc@iarc.fr)	No conflict of interest

2. Proposed update to the Working Procedures of the *IARC Handbooks of Cancer Prevention*

(for reference, see the current Working Procedures for primary prevention and for screening, full text available as PDF at: http://handbooks.iarc.fr/docs/Handbooks-Working_Procedures-Primary_Prevention.pdf

W O R K I N G P R O C E D U R E S - handbooks.iarc.fr

handbooks.iarc.fr

3 The Working Procedures describe theobjectives and scope of IARC Handbooks of 4 Cancer Prevention programme, the scientific principles and procedures used to develop a 5 Handbook, the types of evidence considered, and scientific criteria that guide the 6 evaluations. The Working Procedures should be consulted when reading aHandbook or 7 summary of evaluations made bythe IARC Handbooks.

and http://handbooks.iarc.fr/docs/Handbooks-Working_Procedures-Screening.pdf, respectively)

Location of text to be updated: Document	As requested, here are my comments on working procedures for IARC Handbooks for Cancer Prevention, Screening. A.4. Meeting Participants (1) In describing the criteria for selecting participants for working groups, it is stated that "consideration is also given to demographic and gender diversity and balance of scientific

	findings and views". I understand and agree that gender and demographic diversity is desirable. However, these should be a secondary consideration after it has been ascertained that the person being considered has the necessary expertise and scientific qualifications. I assume that this is IARC policy, but I have seen instances where this policy may not have been rigorously followed. More importantly, I am concerned by the statement that "scientific …views" are considered. Ideally scientists should be totally objective in their assessment of scientific information and should not have pre-formed views. One should beware of persons who advocate for one position or another. It may indicate that they have conflicts of interests or biases for other reasons that influence their judgment. I suggest that this statement be reconsidered. B.4. Efficacy and Effectiveness I strongly disagree with the distinction made between studies of efficacy and studies of effectiveness. As written, it is implied that randomized trials are the only kinds of studies that assess efficacy, and that all observational studies are studies of effectiveness. Consider first randomized trials. In a trial of screening, individuals are randomized to a screening group or a control group. To avoid biased results, analyses are based on intent to screen. That is, endpoints (mortality or incidence of advanced disease) are compared in those randomized to screening or not, regardless of whether those who were randomized to screening actually were screened (or received all of the screenings in studies of more than one round of screening), and regardless of whether some of those in the control group received some screening. (The results of randomized trials can also be influenced by other factors, such as loss to follow-up.) The ratio of the rates of the endpoint in the two groups (e.g. the mortality rate in the screened group divided by the mortality rate in the control group) is the measure of the success of the intervention. As with all studies, it is a measure of the benefit of the screening under the circumstances under which the study was performed. Is this a measure of efficacy or

	effectiveness? The answer is "both." Now consider the observational studies. First, it must be recognized that a screening modality must be efficacious to be effective. Therefore, any study with results which are considered valid and which are interpreted as showing effectiveness is, by definition, evidence for efficacy. Furthermore, some observational studies, at least theoretically, could come closer to measuring the true magnitude of the beneficial effect attributable to a screening modality (i.e. its inherent efficacy) than some controlled trials because they aim to measure the influence of the screening modality in those actually screened, rather than those randomized to screening as in the trials. For example, in a case control study of screening for breast cancer, women who present with advanced breast cancer would be compared with a sample of women without breast cancer who are selected from the same population from which the cases came, and the past history of screening would be ascertained for all of the women. The odds ratio (as an estimate of the relative risk) of the risk of advanced disease in screened vs. unscreened women would then be calculated. This ratio is closer to an estimate of the true impact of the screening on the risk of advanced disease than would be the results of a trial with some women in the screened group not being screened and some women in the control group being screened. This is not to say that case-control studies are better than trials. They are not. But they do provide evidence for efficacy. A similar argument could be made for observational cohort studies. In summary, the studies of efficacy and effectiveness should not be categorized by method. The results of all relevant trials and observational studies should be considered as providing evidence for assessing whether a screening modality is efficacious. Also, the closer that a study comes to measuring the influence of screening among those actually screened the more likely it is to be measuring the potential magnitude of the beneficial effect of the screening modality.

	6. Evaluation. If my argument above is accepted, then the evaluation should be made separately for efficacy and effectiveness. Efficacy is a function of the screening modality itself. I would use the categories of sufficient, limited, inadequate, and evidence against for efficacy only. The criteria for each of the categories seem reasonable to me, with one exception. In the criteria for evidence against, I would eliminate the last criterion, i.e. "There is evidence showing that harms overweight (sic) benefits from the specific intervention." A screening modality may be efficacious and still not be useful because of its harms. I would then separately discuss effectiveness, which is a measure of how well the screening modality works in actual practice. I don't see the value of evaluating effectiveness in the same way as evaluating efficacy, because the effectiveness is so dependent on the local circumstances under which a screening program is conducted. A screening modality can be efficacious and be effective in some situations and not effective in others. Instead, I suggest that the specific factors that are thought to enhance and reduce effectiveness be identified and the evidence for their influence on effectiveness be evaluated. This is where such things as acceptability, compliance, and benefits vs. harms are considered. The likely impact of these factors should be discussed qualitatively. Quantifying their impact is probably not practical or useful because the magnitude of the effect is so dependent on local circumstances. It is also probably not practical to evaluate each factor using the same four categories used in the evaluation of efficacy. I hope that you find my comments useful. Thank you for asking for my input. Sincerely,

	David B. Thomas

Public Comments Form

To Propose an Update to the Working Procedures of the *IARC Handbooks of Cancer Prevention*

Please send completed comment forms to hbworkingproc@iarc.fr.

1. Name and affiliation of commenter

Your name	Robert A. Smith, PhD
Your principal affiliation	American Cancer Society
If another party suggested that you submit this nomination, please identify	Click here to enter text.
WHO Declaration of Interests form (to sign and submit via hbworkingproc@iarc.fr)	No interests to declare; Disclosure Form on file

2. Proposed update to the Working Procedures of the *IARC Handbooks of Cancer Prevention*

(for reference, see the current Working Procedures for primary prevention and for screening, full text available as PDF at: http://handbooks.iarc.fr/docs/Handbooks-Working_Procedures-Primary_Prevention.pdf and http://handbooks.iarc.fr/docs/Handbooks-Working_Procedures-Screening.pdf, respectively)

Comment on IARC Handbooks of Cancer Prevention—Screening—Working Procedures

Page 1—Good introduction

Page 2—lines 3 to 7. I think I would express this differently, and I offer this suggestion to put another suggestion into context. First, the text is a little bit pessimistic. If a cancer screening test is effective and generally has met Wilson and Jungner criteria, then offering it to the public can be endorsed. Benefits will have been judged to outweigh harms over the life course. So, I would suggest the following edits:

Screening requires a commitment among healthy individuals to a series of repeated interactions with health-care providers. It is important for the screening program to reinforce the value of screening and to implement best practices to minimize inconvenience, costs, and the potential for screening fatigue. Effective screening requires an ongoing commitment between the public and health-care providers, and a commitment to use public health resources efficiently.

Page 2, line 25: Suggest "demonstrated," rather than "proven"

Page 2, line 26: Suggest In the past, screening with a given procedure….

Note, there are two circumstances here in which a test would move forward without evidence from an RCT: the first is the example of cervical cancer. The promotion of the Pap test was based on the observation that invasive disease was proceeded by the steady progression of cellular changes. Many years later, there is considerable data suggesting that it is a good screening test, and realistically it was too late to do a RCT (which had not yet been applied to address questions in prevention and early detection research), so careful evaluation of national observational data pushed it over the finish line. A happy ending, but everyone agrees that this is not ideal. The second situation, as we discussed at length during the preparation of the CRC Handbook, is the introduction of alternative screening tests after the efficacy of early detection has been demonstrated earlier with a different test. In this case, "earlier is better" has already been demonstrated.

Page 2, line 32: I think we could be more assertive here…"should be avoided" seems too mild. I know this is all narrative and just a lead up to the methodology, but…..

However, uncontrolled interventions in which individuals are exposed to unknown risks to achieve a benefit that has not yet been proven cannot be supported. To avoid the uptake of "wild type" screening, promising early detection interventions should be promptly evaluated in experimental settings to determine if they are efficacious.

Page 3, lines 1-8: For lines 1-2, I'd suggest being more specific about what we mean by different populations, and perhaps provide some examples. I don't think you're referring to subpopulations, such as the different levels of performance by age

Comment [A1]: Thus far, it always requires repeated interactions. I think you could drop "usually."

Deleted: usually

Deleted: between 3 "healthy" individuals and

Deleted: , which can be inconvenient and 4 costly. Furthermore, effective

Deleted: 5

Deleted: and has inherent public health 6 costs.

Deleted: 30

Deleted: and benefits should be 31 avoided.

within a population, i.e., cervical cancer screening being less effective in older vs. younger women if the same sampling tool is used, or the challenge of greater breast density in younger women in breast cancer screening. Lines 2-5, here again, I suggest that the potential for harms and excess costs should be those that remain after rigorous efforts focused on quality assurance. In the U.S., for example, there is a great deal of complaining about harms, but little effort to reduce avoidable hams. Line 5-8, suggest you be more specific about pertinent health services, i.e., diagnosis and treatment services, and finally, line 8, it often is the case that the what is accomplished by a RCT is the avoidance of bias, but as technology and experience improve, the performance in the community setting exceeds (effectiveness) what was achieved in the trial. This certainly is the case with breast. I think the sentence is fine as is, but you could say "achieve or exceed the outcomes…..

Meeting participants seems OK—I'm increasingly concerned about the way conflicts of interest are being approached, specifically that attention to avoiding COIs has created a kind of a "guilty until proven innocent" environment. In journal disclosures, all relationships are described as "your conflicts of interest." I fully agree that there needs to be full disclosure of all interests, but these should not be described as conflict of interests until they are determined to be real or possible conflicts, including the appearance of a COI ("possible COI" or "the appearance of a COI"s preferable to apparent, since aren't "real" and "apparent" the same?). It also should be possible for a person to have an interest or relationship, or there could be a relationship with the individual's institution and not that individual directly, that could be determined to pose no conflict at all. Shall we just call these interests, and then place emphasis on whether they are determined to be either real or constitute the appearance of a potential COI that would cause some doubt about the credibility of the work? Just a thought.

Review and Evaluation Process

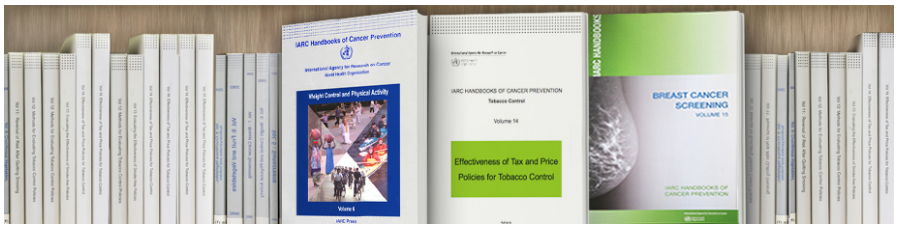
Page 6, line 5—I think everyone would agree that the pace of work when the Working Group meets in Lyon is intense and demanding, and towards the end of the 8-day period, that pace intensifies during the critical review and voting period. The two experiences I had (Breast and Colorectal) would have benefited from at least 1 extra day, and perhaps 2. Less stress during one of the most critical periods of the Handbook R&D.

Page 11, line 5—Perhaps it would be informative to note that estimating overdiagnosis, as an adverse outcome of screening, is extremely difficult, since it depends on measuring excess incidence in the context of screening, comparing an exposed group with an unexposed group. Usually, the data are not so cooperative. The possibility that there is overdiagnosis is an important consideration and the evidence can be described, but thus far, we don't have good estimates for the rate of overdiagnosis for any cancer.

Last comment—I expected to see some discussion on the application of new technology in the presence of previously proven technology….will that be something that is the focus of the upgrade in the methodology?

International Agency for Research on Cancer
World Health Organization

IARC Handbooks of Cancer Prevention



WORKING PROCEDURES

A. GENERAL PRINCIPLES AND PROCEDURES

Back to Table of Contents

1. Background

The global burden of cancer is high and continues to increase: the annual number of new cases was estimated at 14.1 million in 2012 and is expected to reach 22.2 million by 2030 (Ferlay et al., 2015). With current trends in demographics and exposure, the cancer burden has been shifting from high-resource countries to low- and medium-resource countries.

1. Seems low

Prevention of cancer is one of the key objectives of the International Agency for Research on Cancer (IARC). Cancer prevention can be achieved by primary prevention - aimed at preventing the occurrence of cancer - or by secondary prevention - aimed at diagnosing cancer sufficiently early to reduce related mortality and suffering.

2. Chemoprevention as another mode of secondary prevention?

Screening and early clinical diagnosis are the principal instruments of secondary prevention of cancer and a fundamental component of any cancer control strategy. Screening may enable detection of cancer sufficiently early that cure and resulting reduction in mortality and suffering from the disease are realistic possibilities given suitable treatment. Screening for some cancers, such as cervical or colorectal cancer, may also detect precancerous lesions, effective treatment of which can prevent occurrence of cancer.

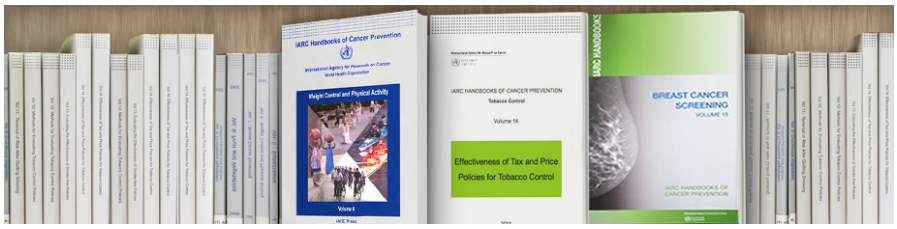
When screening is planned as part of a cancer control programme, only procedures proved to be effective (see below) should be proposed to the general population. Screening usually requires repeated interactions between "healthy" individuals and health-care providers, which can be inconvenient and costly. Furthermore, effective screening requires an ongoing commitment between the public and health-care providers and has inherent public health costs.

Updated 14 November 2017

IARC, 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France - Tel: +33 (0)4 72 73 84 85

International Agency for Research on Cancer
World Health Organization

IARC Handbooks of Cancer Prevention



WORKING PROCEDURES

B. SCIENTIFIC REVIEW AND EVALUATION

Back to Table of Contents

4. Efficacy and effectiveness of a screening procedure

For the evaluation of both efficacy and effectiveness, the Working Group considers the following general principles in making judgements about the available studies:

- Relevance of the study;
- Appropriateness of the study design and analysis to the question being asked;
- Adequacy and completeness of the presentation of the results;
- Degree to which chance, bias, and confounding may have affected the results.

4.1 Efficacy

In this section, evidence from randomized controlled trial (RCT) studies is reviewed. All aspects of study design and analysis are critically discussed. Indicators of the efficacy of the procedure in terms of mortality or incidence, as well as other relevant indicators, such as the detectable phases of the natural history of the disease, are presented.

Aspects that are particularly important in evaluating RCTs are: the selection of participants, the nature and adequacy of the randomization procedure, evidence that randomization achieved an adequate balance between the groups, exclusion criteria used before and after randomization, compliance with the intervention in the screened group, and â€œcontaminationâ€ of the control. Other considerations include the means by which the outcome (preneoplastic lesions or cancer) was determined and validated (either by screening or by other means of detection of the disease), the length and completeness of follow-up of the groups, and the adequacy of the analysis.

3.

When RCTs are lacking, efficacy cannot be directly evaluated, but only indirectly inferred from observational studies (see below).

4.2 Effectiveness of population-based screening

The impact of the screening procedure when implemented in defined populations is examined in this section.

In this section, mostly observational studies are reviewed, conducted in

o The screening procedure is consistently associated with no reduction in mortality from or incidence of invasive cancer, and chance, bias, and confounding can be ruled out with reasonable confidence.

o There is evidence showing that harms overweight benefits from the specific intervention.

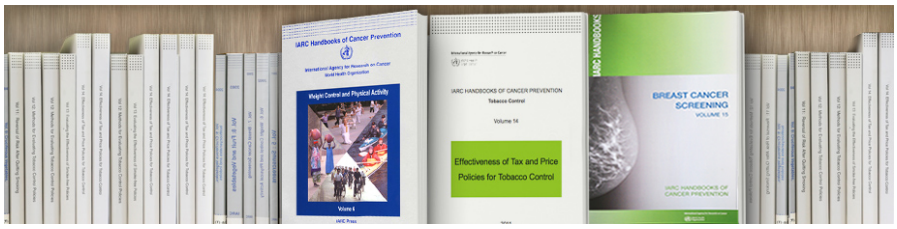
4. ?

Updated 14 November 2017

IARC, 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France - Tel: +33 (0)4 72 73 84 85

International Agency for Research on Cancer
World Health Organization

IARC Handbooks of Cancer Prevention



WORKING PROCEDURES

A. GENERAL PRINCIPLES AND PROCEDURES

Back to Table of Contents

1. Background

Prevention of cancer is one of the key objectives of the International Agency for Research on Cancer (IARC). The aim of the *IARC Handbooks of Cancer Prevention* series is to review and evaluate scientific information on interventions that may reduce the incidence of or mortality from cancer. As a result of The *Handbooks* evaluations, national and international health agencies have been able, on scientific grounds, to take measures to develop interventions or recommendations that will reduce the risk of developing cancer.

The criteria guiding the evaluations were first established in 1995 at inception of the *IARC Handbooks* series, and were revised in subsequent volumes.

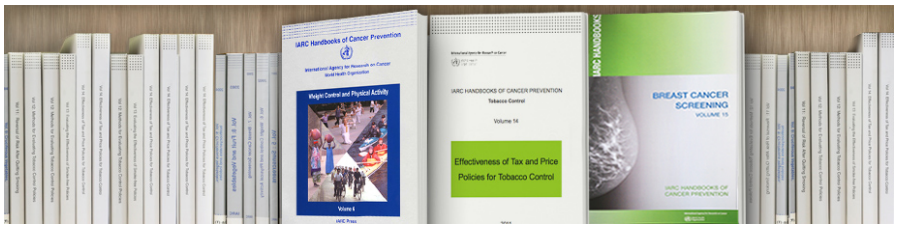
5.

Posted 5 July 2016

IARC, 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France - Tel: +33 (0)4 72 73 84 85

International Agency for Research on Cancer
World Health Organization

IARC Handbooks of Cancer Prevention



WORKING PROCEDURES

A. GENERAL PRINCIPLES AND PROCEDURES

Back to Table of Contents

2. Objective and scope

The objective of the *IARC Handbooks* programme is the preparation of critical reviews and evaluations of the evidence that a particular intervention can prevent cancer. The evaluations, which are prepared by a Working Group of international experts, are scientific judgements about the available evidence on efficacy, effectiveness, and safety of a wide range of cancer-preventive interventions. No recommendation is given with regard to national or international regulations or legislation, which are the responsibility of individual governments and/or other international authorities. The *IARC Handbooks* may assist national and international authorities in devising programmes of health promotion and cancer prevention, and in making benefit-risk assessments.

In this document, the term "intervention" refers to any chemical, activity, or strategy that is subject to evaluation in a *Handbook*. Cancer-preventive interventions encompass pharmacological, immunological, dietary, and behavioural interventions that may delay, block, or reverse carcinogenic processes, or reduce underlying risk factors.

Preventive interventions can be applied across a continuum of: (1) the general population; (2) subgroups with particular predisposing host or environmental risk factors, including genetic susceptibility to cancer; (3) persons with precancerous lesions; and (4) cancer patients at risk of developing second primary tumours. Use of the same interventions in the treatment of cancer patients to control the growth, metastasis, and recurrence of tumours is considered to be patient management and not prevention, although data from clinical trials of such interventions may be pertinent when reaching anevaluation.

6. Separate these two out

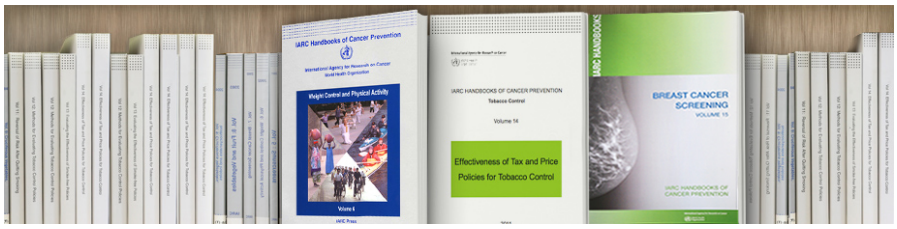
7.

Posted 5 July 2016

IARC, 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France - Tel: +33 (0)4 72 73 84 85

International Agency for Research on Cancer
World Health Organization

IARC Handbooks of Cancer Prevention



WORKING PROCEDURES

B. SCIENTIFIC REVIEW AND EVALUATION

Back to Table of Contents

2. Studies of cancer prevention in humans

This section includes all pertinent experimental and observational epidemiological studies of cancer prevention in humans, with cancer as an outcome (see Part A, Section 4). Studies of biomarkers as indicators of the intervention are included in Section 4 when they are relevant to an evaluation of the cancer-preventive effect in humans.

(a) Types of study considered

This section focuses on studies that assess the prevention of cancer as an outcome in humans. Relevant evidence is normally provided by experimental studies (for example, randomized clinical trials and community intervention trials), and analytical observational studies, primarily cohort studies and case-control studies. For certain interventions applied at the population level, well-designed ecological studies (studies measuring both outcome and exposure on the aggregate, or population, level) or interrupted time-series studies may also be informative. Cross-sectional studies, descriptive epidemiological studies, case-series and case reports are usually not reviewed. The uncertainties that surround the interpretation of such studies make them inadequate, except in exceptional circumstances, to form the basis for inferring a preventive relationship. However, when considered together with experimental and analytical observational studies, these types of study can sometimes contribute to the decision of the Working Group as to whether or not a causal relationship exists. 8. ?

Intervention studies are experimental in design - that is, the use of, or exposure to, the intervention is assigned by the investigator. Experimental studies can provide the strongest and most direct evidence of a protective or preventive effect; however, the use of such studies is limited for practical and ethical reasons and the subjects are often drawn from select groups that may not represent the population at large.

In exceptional cases, epidemiological studies on advanced preneoplastic lesions and other end-points thought to be relevant to cancer are also reviewed in this section. The results of such studies may strengthen inferences drawn from other studies.

(b) Quality of studies considered

In considering whether a particular study should contribute to the evaluation of an intervention, the Working Group considers the following

suppression of effects that are on the pathway to cancer. The mechanistic evidence can be strengthened by findings of consistent results in different experimental designs, by the demonstration of biological plausibility, and by coherence of the overall database.

The Working Group considers whether multiple mechanisms might contribute to cancer prevention, whether different mechanisms might operate in different dose ranges or at different sites, or whether separate mechanisms might operate in a susceptible group.

For complex interventions, such as food categories, the chemical composition and the potential contribution of different nutrients known to be present may be considered by the Working Group in its overall evaluation of cancer prevention.

(d) Overall evaluation

Finally, the body of evidence is considered as a whole, and summary statements are made that encompass the effects of the intervention with regard to cancer-preventive effects in humans. The overall evaluation is described according to the wording of one of the following standard categories. The categorization of an intervention is a matter of scientific judgement that reflects the strength of the evidence derived from studies in humans and in experimental animals, and from mechanistic and other relevant data.

(i) The intervention prevents cancer (Group A)

This category is used for interventions for which there is *sufficient evidence* of a cancer-preventive effect in humans.
The sites on which the evidence in humans is based are given.

(ii) The intervention probably prevents cancer (Group B1)

This category is used for interventions for which there is *limited evidence* of a cancer-preventive effect in humans and *sufficient evidence* in animals. An intervention may also be classified in this category when there is *limited evidence* in humans, less than *sufficient evidence* in experimental animals, and strong supporting evidence from mechanistic and other relevant data that the mechanism(s) of prevention also operates in humans.

The sites on which the evidence in humans is based are given.

(iii) The intervention possibly prevents cancer (Group B2)

This category is used for interventions for which there is *inadequate evidence* in humans, and *sufficient evidence* in experimental animals. An intervention may also be classified in this category when there is *inadequate evidence* in humans, *limited evidence* in experimental animals, and strong supporting evidence from mechanistic and other relevant data that the mechanism(s) of prevention also operates in humans.

(iv)The intervention is unclassifiable as to its cancer-preventive effects (Group C)

This category is used for interventions for which the evidence is *inadequate* in humans and less than *sufficient* in experimental animals. Interventions that do not fall into any other group are also placed in this category.

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

This category is used for interventions for which there is *evidence suggesting lack of a cancer-preventive effect* both in humans and in experimental animals.

(e) Rationale

The reasoning that the Working Group used to reach its evaluation is presented and discussed. This section integrates the major findings from studies in humans, studies in experimental animals, and mechanistic and other relevant data. It includes concise statements of the principal line(s) of argument that emerged, the conclusions of the Working Group on the strength of the evidence for each group of studies, and an explanation of the reasoning of the Working Group in weighing data and making evaluations. The human populations that were the subject of study should be identified. Additionally, important health concerns identifiedâ€"such as adverse effects, including cancer-causing properties should be clearly addressed.

10.

When there are significant differences in scientific interpretation among Working Group Members, a brief summary of the alternative interpretations is provided, together with their scientific rationale and an indication of the relative degree of support for each alternative.

Posted 5 July 2016

IARC, 150 Cours Albert Thomas, 69372 Lyon CEDEX 08, France - Tel: +33 (0)4 72 73 84 85

Public Comments Form

To Propose an Update to the Working Procedures of the *IARC Handbooks of Cancer Prevention*

Please send completed comment forms to hbworkingproc@iarc.fr.

1. Name and affiliation of commenter

Your name	Ludovic Reveiz
Your principal affiliation	Pan American Health Organization
If another party suggested that you submit this nomination, please identify	Click here to enter text.
WHO Declaration of Interests form (to sign and submit via hbworkingproc@iarc.fr)	PAHO / WHO staff

2. Proposed update to the Working Procedures of the *IARC Handbooks of Cancer Prevention*

(for reference, see the current Working Procedures for primary prevention and for screening, full text available as PDF at: http://handbooks.iarc.fr/docs/Handbooks-Working_Procedures-Primary_Prevention.pdf and http://handbooks.iarc.fr/docs/Handbooks-Working_Procedures-Screening.pdf, respectively)

Location of text to be updated: Document Section Page number Line number	Line 25 to 28 IARC Handbooks of Cancer Prevention Screening WORKING PROCEDURES **5. Review and evaluation process** 5 18 - 29
Current text	"IARC performs literature searches to compile the relevant bibliography in relation to the topic that will be evaluated. Meeting participants are expected to supplement the IARC literature searches with their own searches of published evidence".
Proposed update (revised text)	"IARC will develop the Handbooks of Cancer - Prevention Screening in accordance with standard procedures set out in the World Health Organization handbook for guideline development [ref]. The quality of the evidence supporting the recommendations will be graded using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) [ref] and the Confidence in the Evidence from Reviews of Qualitative Research (CERQual) approaches [ref]. The GRADE Evidence to Decision Framework (EtD) approach incorporates intervention effects, resources, values, equity, feasibility and acceptability criteria among others".
Brief rationale for update (max. 200 words)	By incorporating the WHO, GRADE and CERQual approaches, other sections should be modified in the SCIENTIFIC REVIEW AND EVALUATION section. Main points for suggesting the modification are: 1. WHO/GRC, GRADE and CERQual are explicit, standardize, comprehensive and widely recognized approaches for developing health evidence-informed recommendations 2. The current method will not necessarily summarize or cite the entire body of literature on the intervention being evaluated (Page 7, Line 6-8). There are no clear criteria for deciding this.

	GRADE provides a comprehensive and standardize approach for decision making. 3. More methodological details are needed in the document (how systematic reviews will be conducted, how the quality of a body of evidence on the effectiveness will be judged and rated and which Strength of the Recommendations Grading System will be used). GRADE has developed an explicit approach for diagnostic tests and strategies that can be applied for the development of the handbooks of cancer prevention - screening. 4. Many countries have incorporated WHO/GRC and GRADE approaches for the development of their national guidelines. Having a similar approach could facilitate the adaptation / adoption of the IARC handbooks of cancer prevention – screening by Member States. PAHO is currently developing and international database of GRADE guidelines (BIGG). Although still under construction, we have identified so far more than 700 GRADE guidelines published in the last five years worldwide (only half are currently available in the database). 5. Using this approach could facilitate subsequent updates by IARC or MS
References, if any (max. 5)	1. WHO handbook for guideline development – 2nd ed https://www.who.int/publications/guidelines/guidelines_review_committee/en/ 2. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. https://gdt.gradepro.org/app/handbook/handbook.html 3. Pan American Health Organization 2018. Strengthening national evidence-informed guideline programs. A tool for adapting and implementing guidelines in the Americas ISBN: 978-92-75-12016-3 http://iris.paho.org/xmlui/handle/123456789/49145

Comments

Section A2.
Line 56. The PPV of many recommended screening tests is well below 50%. Therefore, the statement
"Screening tests sort out apparently-well people who probably have a disease from those who probably do not".
is not accurate. It is more accurate to say that screening separates out people who are more likely to have the disease from those who are less likely.

Line 95 – Since screening generally does not "prevent" cancer, suggest changing the wording of this sentence
" … the Handbooks base their evaluation of the strength of the evidence that a putative preventive intervention actually can prevent cancer on the quality and results of all relevant research".
Suggest something like "reduce the burden of cancer" or "prevent cancer incidence or mortality".

Section A3.
Line 105-
It is not clear what is meant by
"The intervention is of putative protective value, but the efficacy, effectiveness or safety have not been established formally"
For example, screening mammography and screening for colorectal cancer were both subjects of handbooks and at the time of the handbook, both had had many trials showing efficacy and were recommended by many entities. What does "established formally" mean?

Section B4.
Line 333 - The term "epidemiologic studies" (in the heading) is often thought to refer only to observational studies, as opposed to randomized trials. Cost-effectiveness is also commonly not considered an "epidemiologic study". Suggest another term, for example, "Experimental and observational studies of each screening method".

Section B4.1
In the discussion of experimental studies, or randomized trials, somewhere the distinction between ITT (intent-to-treat or intent-to-screen) and per-protocol estimates of the RR (or other metric) should be discussed. Some consider ITT estimates to be more relevant for judging effectiveness and per-protocol estimates more relevant for judging efficacy, though not all agree. Any per-protocol estimates must be based on methods that avoid selection bias.
Line 407 –

The sentence “Confidence intervals … that could be produced by chance alone” is confusing. For a point estimate, say RR, whose 95% CI does not cross 1.0, the general conclusion would be that the protective effect is not due to chance; therefore, saying the 95% CI is what could be produced by chance alone is not accurate or at least misleading. Could say the “range of plausible values”.

Section B4.1d
Line 523
It states

“At least, however, (i) a cross-sectional comparison of new test’s accuracy with that of a screening test that has been established to prevent cancer death and (ii) a randomised controlled trial to establish whether, in comparison with the old technology, the new technology can reduce risk of interval cancer should be performed”

The first study (cross-sectional comparison) makes sense. However, the 2nd (trial for interval cancers) is problematic. Interval cancers are generally fairly rare and the sample size for such a trial would frequently have to be enormous and thus not practical. Note that the current TMIST trial (comparing 2D mammography with tomosynthesis) does not have as its endpoint the interval cancer rate, but instead a composite endpoint of advanced disease. There should also be some consideration here of how similar mechanistically the new technology is to the established one (e.g., colonoscopy vs sigmoidoscopy), with less direct evidence needed for the new technology the closer that technology is mechanistically to the established one.

Section B4.2
Line 586 – Interval cancers are not a harm of screening, but rather a failure of the efficacy of screening. That is very different. An interval cancer is one diagnosed by symptoms in a screening setting; in the absence of screening, essentially all cancers would be diagnosed that way. The only way an interval cancer (or a false negative) would be a harm is if knowledge of the prior negative screen delayed diagnostic work-up. That is speculative and to my knowledge, there is no real evidence for it. It is certainly not a major harm of screening, in general.
Line 586 – In addition to harms of “adverse consequences of unnecessary treatment of an overdiagnosed cancer”, there is harms of overdiagnosis per se (being labelled as having cancer and psychological harms of being told one has cancer), as well as harms of surveillance, when treatment is delayed.
Somewhere in this section, the concept of ancillary findings should be discussed. These are a potential harm, and also a potential benefit, e.g., with low-dose CT lung cancer screening or CT colonography.

Section B4.3, 4.4.
It should be emphasized that benefits-to-harm ratios and cost-effectiveness are very dependent on disease prevalence in a population to be screened, in that costs and harms are generally relatively independent of disease prevalence (except for overdiagnosis) while benefits are often directly proportional to prevalence. Often such data (on harms and costs) are from studies conducted in developed countries where the

prevalence of the cancer may be substantially higher than in less developed countries. Therefore, extrapolating benefits-to-harms ratios and cost-effectiveness ratios to other settings must be done taking into account disease prevalence. This is in addition to the fact that costs of medical tests and treatments may clearly differ between these settings, as is pointed out in the current text.

Section B6.1.
For the evaluation for Group A of the benefits and harms and whether the benefits outweigh the harms, it should be noted that this depends on the population setting, and specifically on underlying disease prevalence. Benefits may outweigh harms in high prevalence settings (including where the studies were conducted) but not in lower prevalence settings.